CUK-H05001-48-P009474

ISSN 0001 253X
Volume 51
Number 1
January 1999

48

27/1999
LOCL

# Aslib Proceedings

## New information perspectives

Ranganathan and the Net: using facet analysis to search and organise the World Wide Web
David Ellis and Ana Vasconcelos

Send a boy – or dial it yourself? Numbering for the information society
Ralph Adam

Football information services: fanzines, 'Match of the Day' and the modem
Richard Green

The News International Newspapers' intranet
Gertrud Erbach and Lynda Iley

Aslib THE ASSOCIATION FOR INFORMATION MANAGEMENT

*Aslib Proceedings* is a long-standing professional journal of international repute, which has a large and wide subscriber base. The journal brings currency, authority and accessibility to the reporting of current research, issues and debates in the broad area of information work. Above all the journal wishes to provide research and comment in a form that is easily and quickly understood: a fresh, rigorous, but unfussy, writing style is what is aimed for. The needs of busy practitioners (and academics) is very much in mind. Articles are now refereed to meet the growing need for accredited and authoritative information. To ensure that this does not result in an unacceptable loss of currency, articles are speedily refereed by the journal's large, eminent and multi-disciplinary Editorial Board (assembled for just such a purpose). In exceptional cases they may be referred to an external panel of consultants. The journal's monthly publishing pattern helps in maintaining the currency of reporting. In consequence the submission of research in progress articles is encouraged.

The boundaries of information work are rapidly changing and widening. *Aslib Proceedings* has a role to play in the repositioning of the information profession by encouraging submissions from information managers, providers, packagers and consumers of all kinds – and especially those from the related information fields of journalism, electronic publishing, communication and Internet studies. An interdisciplinary Editorial Board has been assembled to assist in the process.

Prospective authors should submit articles of 2,000-6,000 words in length. Articles should be accompanied by a 100-150 word informative abstract and the numerical form of referencing should be adopted. In style and tone, submissions should fall between that of the journal's sister publications: *Managing Information* and *Journal of Documentation*. Articles should be furnished by email or on disc to David Nicholas, Editor, Aslib Proceedings, Department of Information Science, City University, Northampton Square, London EC1V 0HB. Email: nicky@is.city.ac.uk. The Editor is happy to provide advice on an informal basis to potential authors.

**Editorial Board**
Dr David Nicholas, *Head of Department, Department of Information Science, City University;* John Akeroyd, *Director, Learning Resources, South Bank University;* Sarah Blair, *Head of Publications, Aslib;* Peter Chapman, *Information Manager, NewQuest North East;* Professor Peter Cole, *Head of Department, School of Journalism, University of Central Lancashire;* Dr Tom Dobrowolski, *Senior Lecturer, Institute of Information Science and the Book, University of Warsaw;* Harry East, *Senior Research Fellow, Department of Information Science, City University;* Mike Hyman, *Managing Director, Head Software International;* Robin Hunt, *Creative Director, AreHause;* Nat Lievesley, *Network Manager, Centre for Policy on Ageing;* Helen Martin, *Information Manager, Research and Information Unit, The Guardian*; Dr Ian Rowlands, *Lecturer, Department of Information Science, City University;* Richard Withey, *Director, New Media, News International.*

**Editorial Team**
*Editor:* David Nicholas
*Deputy Editor:* Ian Rowlands
*Editorial Manager:* Kay Nicholas
*Editorial Assistants:* Peter Williams (City University), Colette Coles (City University)

010.5
AS 52 PROC

# CONTENTS

3-10 Ranganathan and the Net: using facet analysis to search and organise the World Wide Web
David Ellis and Ana Vasconcelos

11-19 Send a boy – or dial it yourself? Numbering for the information society
Ralph Adam

20-29 Football information services: fanzines, 'Match of the Day' and the modem
Richard Green

30-34 The News International Newspapers' intranet
Gertrud Erbach and Lynda Iley

p 9474

# Ranganathan and the Net: using facet analysis to search and organise the World Wide Web

*David Ellis*
*Department of Information Studies, University of Sheffield, Sheffield, S10 2TN*
*Email: D.Ellis@sheffield.ac.uk*

*Ana Vasconcelos*
*School of Computing and Management Sciences, Sheffield Hallam University, Sheffield, S1 1WB*

**Abstract** *This paper documents the continuing relevance of facet analysis as a technique for searching and organising World Wide Web based materials. The two approaches underlying WWW searching and indexing – word and concept based indexing – are outlined. It is argued that facet analysis as an a posteori approach to classification using words from the subject field as the concept terms in the classification derived represents an excellent approach to searching and organising the results of WWW searches using either search engines or search directories. Finally, it is argued that the underlying philosophy of facet analysis is better suited to the disparate nature of WWW resources and searchers than the assumptions of contemporary IR research.*

## Introduction: the faceted approach

Ingwersen and Wormell[1] have already pointed out the continuing relevance of facet analysis for contemporary information retrieval. The intention here is to indicate the relevance of facet analysis in relation to searching and organising material on the World Wide Web. This is of relevance not only for those who design or develop web pages but also for anyone searching the web. Facet analysis was devised by the Indian mathematician and documentalist S. R. Ranganathan[2] and is essentially a concept based approach to classification, which involves analysis of complex subjects into groups of simple concept terms, the groups being referred to as facets. Compound subject descriptions are then synthesised from combinations of the simple concept terms. For this reason faceted classification schemes are also often referred to as analytico-synthetic schemes. There are very few pure examples of such schemes in practical use (BLISS is the exception) but the principles of facet analysis, particularly the use of synthesis, have been widely influential and have been incorporated in, or used as the basis of revision of, otherwise largely enumerative classification schemes such as the Dewey Decimal Classification and the Universal Decimal Classification Scheme (UDC).

The faceted approach proceeds by analysis of a restricted subject domain into simple concept terms; grouping of the simple concept terms into facets; devising a notation to maintain the simple concept terms in order in the facets; and then synthesising document descriptions by the combination of simple concept terms from the various facets. The result is a controlled vocabulary of simple concept terms organised into facets for pre-coordinate or post-coordinate operation. It is possible to combine any individual concept or term with any other which provides flexibility for both the indexer and the searcher. The subject chosen should be relatively specific or narrow. Facet analysis is best seen as a tool for developing specialist classification schemes rather than for generating over-arching general classifications although some work has been done in this latter area. The concept terms should be derived from and reflect the actual terms used in the field rather than being based on some *a priori* conceptualisation of the subject. Therefore, the first step in developing a facet classificaton is to compile a small database of terms from primary, secondary or tertiary sources such as dictionaries, encyclopaedias, standard texts, articles or other authoritative sources.

Following initial analysis of these terms, a more thorough examination of the subject can be used to compile a more comprehensive lists of terms. These terms will then be analysed into facets by the application of different principles of division. From this analysis the concept terms are grouped into facets corresponding to the principles of division applied. The application of a principle of division within a facet leads to the creation of a sub-facet. The order in array or order of terms within facets and sub-facets should be decided in terms of usefulness or naturalness to the user. There is no hard and fast rule for this, often some natural, temporal or hierarchical order will be present and this can be used. If no such order suggests itself then alphabetical or numerical order is usually adopted.

### The Internet and the World Wide Web

The Internet has its origins in a network, called the ARPAnet, developed in the early 1970s by the Advanced Research Projects Agency (ARPA) in the US. It was, at first, used for the transfer of intelligence within the US Department of Defence, but, later on, it was adopted by the academic community to build an academic network for information exchange[3,4]. By the late 1980s, access to the Net was open to the general public, as network technology introduced client server architectures and ethernet local area networks, supporting end-user access to networks.

By the mid 1990s, the Internet was made of over 60,000 networks, with a growth rate of 10% per month and an estimate of 50 million people worldwide having access to the Internet on email, displaying a growth rate of 1000% per year of use of file search and retrieval tools[3]. The Internet is a voluntary and co-operative effort, where the over 60,000 participating networks agree on the common protocols and rules. Nobody, therefore, is officially in charge of or controls the Internet.

The World Wide Web (or the WWW or the W3) is the multimedia part of the Internet, displaying a hypertext type of structure and search facilities (browsing and navigation). It was first developed for sharing documents between nuclear physicists of the European Particle Physics Laboratory (CERN) in Switzerland, 1989, but the first commercial web software, by NeXT, in 1991, soon popularised this form of access to the Internet.

The main characteristics of the WWW organisation and structure are:

- organising documents into pieces of information (pages), using a set of rules which tag and format the documents, the Hypertext Markup Language (HTML);
- every individual document or page is assigned a unique address, called the Uniform Resource Locator (URL);
- each URL can be linked with a hypertext type of link to other URLs and even pieces of information within each document (buttons) can be linked to URLs or to other pieces of information;
- these documents can be searched through interactive interface programs which allow users to browse and navigate through the documents and are called web browsers;
- the communication between the web browsers and the web servers is regulated by a common language using a standardised set of rules called Hypertext Transfer Protocol (HTTP)[3,4]; the HTTP allows the interpretation of the HTML signs within each web page, in order to display correctly the page and to enable the transfer of files.

### Issues raised by the Internet and the WWW for ISR

The huge size of the Internet and its growth rate, together with the fact that there is no control over it, mean that:

- a huge volume of information is generated every minute;
- no order or rules are imposed on the generation, distribution, access and use of this information;
- no fully comprehensive record of the different documents is available at the moment;
- no classification and description framework for storing and retrieving these documents has been commonly accepted and established, although there are several classifications which have been applied to parts of the WWW;
- therefore, documents of all kind of format (text files, sounds, images, video clips), type

of information (electronic journals, tourist information, marketing services, commercial sales, library catalogues) and subject (astrology, cookery recipes, satellite images of the Earth, university course guides, government reports, business databases) can be found.

One of the most widely used formal devices for organising knowledge for information retrieval is by indexing. Indexing systems have often been characterised as reflections of the way the human mind organises and uses knowledge. According to Anderson[5]:

> 'Indexing systems represent attempts to extend the organising capabilities of the human mind to [...] artificial (humanly devised) information storage and communication systems [i.e., written, sound and visual documents, transmitted by post, radio, TV, electronic networks and other communication systems]. Indexing systems turn storage and communication systems into information retrieval systems analogous, on widely varying degrees, to the human information retrieval system of the mind.'

Indexing systems have been originated with early collections of systems and their function is essentially to indicate 'the content and related features of a document'[6], with the double purpose of: 'identifying documents on specific subjects and identifying documents on related subjects'[7].

It is therefore not surprising that indexing systems have historically been linked with the theory of the organisation of knowledge and share with it common assumptions[8]:

- human knowledge can be organised through a conceptual framework;
- this conceptual framework is a hierarchy of classes and subclasses;
- classes are formed by grouping things that share the same attributes and characteristics (association – for example, cats and dogs have fur) and by separating things that do not share the same attributes and characteristics (discrimination – for example, dogs have fur, fish have scales).

According to Anderson[9] indexing systems usually comprise the following components:

- a collection of documents (such as the web sites on the WWW);
- a separate set of surrogates (such as titles and summaries of web sites in subject directories like Yahoo! and the WWW Virtual Library or the copies of web pages in search engine databases) which represent these documents;
- and a separate index (also called indexing language) which links the document collection and its surrogates (such as the subject indexes you have used in Yahoo! and in the WWW Virtual Library).

## Indexing and search tools on the WWW: search directories and search engines

The hypertext structure of the web means that retrieval is done through following the links between different web pages, through browsing and navigation. Because each web page will have multiple links, there will be multiple paths to the same piece of information. This way of searching through association has advantages and disadvantages[10].

- The main advantage is that it can lead to finding information people are not aware of, just by following links.
- The main disadvantage is that people usually have to follow different layers of linked information in order to retrieve one specific page or piece of information.
- With the huge increase of the WWW – which doubles its size every four months[11] – the problems associated with this way of searching for information are increased, since there is no comprehensive framework which guides the organisation of the web. As stressed by Tyner[12], searching for a specific piece of information through the over 150 million pages on the web without a central catalogue becomes an almost impossible task.
- Although, the WWW does not have a central catalogue or index, different search tools have been devised in order to assist people in finding information on the largest repository of documents in the World. These search tools can be divided into two main categories: subject directories and search engines.

### *Subject directories*

Subject directories (also called subject trees or subject guides), allow people to browse information by subject such as Accounting and Finance, Astrology, Biology, Health. They are hierarchically organised indexes of different subjects, with links to different web sites on each subject. The searcher can browse through the index in search of relevant subjects and navigate to the relevant web sites, by clicking on hotspots which represent those sites. Subject directories are built by human indexers using indexing and abstracting techniques. Editors constantly review sites, index them and build short descriptions of these sites. Alternatively, web site producers submit their own description of their sites to the organisations which compile these indexes.

These short descriptions are maintained in the subject directory database and linked to the main index. The descriptions usually provide a link to the Home page of a web site, rather than a link to each individual page in a site. The fact that the links are made to the top level of web sites means that subject directories are better adjusted to searching general subjects. Because they are humanly indexed, subject directories tend to retrieve more relevant information than other search tools. However, their drawback is that they take some time to include new sites in their indexes and in reviewing these new sites. Examples of subject directories include: Galaxy *(http://www.einet. net/galaxy.html);* Magellan *(http://www.mckinley.com)* and, Yahoo! *(http://www.yahoo.com)* probably the largest directory, due to using automated programs to search for new sites, although using human indexers to index them[3,10,11].

Apart from general subject directories, such as the above mentioned, there are two other types of directories: specialised subject directories and clearinghouses. Specialised subject directories, as their name indicates, specialise in specific areas and provide access to the most important sites and resources in those areas. There are specialised subject guides in almost all subject areas. Other search directories, known as clearinghouses, are collections of various specialised search directories, either by including the specialised search directory on their own web sites or by providing a link to the web sites of these specialised directories[12].

***Search engines***

Search engines, rather than searching documents through an index, are based on allowing users to enter keywords that are matched against a database. Unlike subject directories, which use human indexers to build their indexes, search engines use software programs which create automatically their own databases, containing lists of web pages. Search engines are composed of three different parts: a program called a spider (or robot or crawler), a database with an index and search software[13,14,15,16,17]. Spiders wander through the web, crawling from site to site, following links between pages. Different search engines use different types of spiders: whereas some visit every possible site they visit, others are based on more selective principles and visit only the most popular sites. The first type of spider finds a huge volume of information in a short space of time, the latter type generates a smaller number of pages, with perhaps more relevant results. Every page found by a spider is stored in a database and an index of its contents is built, by extracting automatically words from the web pages and ranking them alphabetically, using the same principles which inverted files use. The index is, therefore, a list of every word found (except irrelevant words such as *and*, *the*, *an*, called stop words), with a pointer to its location on the database.

Again, different search engines will follow different principles: some will index every single word on every web page the spider found (excluding stop words), other search engines index only the title and top level phrases of a web site. The third element of the search engine is the search software, which is a program which compares search queries keyed in by people (for example, Search: 'marketing companies') with the index, finds matches and ranks them in order of relevance. The criteria for judging relevance will vary according to the search engine[18].

The different approaches which search engines use in crawling the web, finding new pages and indexing them will produce totally different results. This is why when searching the same topic in different search engines there can be very striking differences in the results. Also, because of the way they operate, search engines are more oriented to find larger volumes of information and more specific information than subject directories, because they are based on searching web pages (and not just the top level of sites, as happens with subject indexes) and indexing automatically precise words from the sites (rather than using a pre-defined index) as subject directories do.

Examples of search engines include: AltaVista *(http://altavista.digital.com/);* Excite *(http://www.excite.com/);* HotBot *(http://www.hotbot.com/);* Infoseek *(http://www.infoseek.com/);* and Lycos *(http://www.lycos.com/).* As different search engines have different strengths, new search tools have been created more recently, which allow people to search simultaneously the different databases of search engines, while using one single interface. These tools are called multi-threaded search engines and although they usually do not have all the search flexibility of individual search engines, they are very fast and can search through vast amounts of information. Examples of multi-threaded search engines include: Dogpile *(http://www.dogpile.com)* and Metacrawler *(http://www.metacrawler.com).*

## Word and concept indexing on the World Wide Web

The different approaches used by subject directories and search engines in terms of indexing are not new. Subject directories use a conceptual approach, which people have been using over centuries, which is at the basis of the classification schemes used in many fields (from Library classifications, such as Dewey, to taxonomic classifications in Biology). Search engines use more recent methods, based on the automatic extraction of words into an index, developed and adopted since the first text retrieval software applications in the 1960's and materialised in text retrieval systems based on inverted files. There are, therefore, two main types of methods for indexing or searching for documents on the World Wide Web – word based and concept based.

Word based methods concern the automatic input of words which are recorded in the documents to be described, without any further input or consideration of the exact meaning of each of the terms extracted from the source document. These words will then be used to describe the document. This method is computer based and the words are extracted regardless of their meaning; this is the fastest and more economic approach and it is increasingly used. The search engines on the WWW work through automatic word indexing of web sites[19]. Automatic indexing systems very frequently include statistical devices to assign weights to each term and therefore determine which are the most important terms in the document, by frequency of occurrence. It is then possible to select for representation only the most frequently occurring terms.

Concept based methods, contrary to word based methods, require the identification of the concepts which are represented by the terms used in the documents, rather than just extracting the terms used in the document. Concept based methods use concepts identified by a human indexer in the document, who then selects which are the most appropriate to represent the document. As stressed by Buchanan[20], 'concepts are ideas of things, to be distinguished from names of things'. Because of this, the index terms used to represent the document may often be different than the terms actually used in the document, as the same concept may be expressed through different terms. The creation of concepts in subject indexing is analogous to the creation of concepts in the mind, as it is based on the association of representations of objects, ideas, processes and other entities with related data which pre-exists in the mind, as outlined by Anderson[9] and Foskett[21].

Concept based methods require the intervention of the human mind and imply the analysis of the concepts present on the text and the choice of terms best representing these documents. This method of representation introduces an element of interpretation and is more time consuming and less economic than the previous, but it is focused on the meaning of each term; the search directories in the WWW have a concept based organisation and use human indexing of the different web sites. Concept based indexes are often structured in order to represent the relationships between the different concepts, either through the use of references (for example, documents see also reports) or by grouping related subjects together (for example, religion, Buddhism, Christianity, Hinduism, Islamism, Judaism).

Word indexing, although fast and economic, has problems associated with it, as it does not consider the meaning of words. These problems occur in retrieval, as systems which are word indexed cannot distinguish between words which are spelt the same way but mean something different (a *hard exam* and a *computer hard disk*, for example). Also, different words referring to the same concept (*heart attack* and *cardiac arrest*) cannot be retrieved unless you key in every word form (if you entered the query *'heart diseases'*, you would not retrieve a document which had used the expression *'cardiac diseases'* instead).

Concept based indexing, used by subject directories, tries to address these problems by basing indexing, not on extracting words from the text, but on identifying the meaning of different words in the context they are used and choosing one keyword or expression for each concept. These keywords or expressions will form the subject index of subject directories. In these cases, indexing is carried out by humans. Some search engines (Excite, Euroferret), however, use very sophisticated computer programs, based on statistical and probabilistic calculations and/or on artificial intelligence, to try to simulate the way humans identify concepts[17,22,23]. These programs analyse the co-occurrence of different words which relate to the same subject in order to determine what subject(s) the document is about – for example, the word *hard*:

- associated with other words like *surface*, *stone*, would refer to the physical property of an object being solid;
- associated with words like *exam*, *text*, *experience*, would refer to the degree of difficulty involved in certain situations;
- associated with words like *person*, *man*, *woman*, would refer to personal characteristics, unfriendliness or severity.

## Using facet analysis to search and organise the WWW resources

It is easy to overlook the relevance of techniques such as facet analysis in the context of the accelerating changes in the information environment with first the IT and then the Internet revolutions. The LIS professions, which have their intellectual and conceptual foundations in originally non-IT applications of information management, should recall that the techniques of knowledge organisation which were developed in the context of manual systems and hard copy information sources are not restricted in their application to the historical source of their development. As the use of facet analysis in the creation of thesauri for online databases attests, however, the development of the WWW and the proliferation of WWW information sources has taken place largely outside the sphere of LIS professions or practice. This is not to say that such professions and practices could not serve to benefit the development of the WWW, as already demonstrated by the subject directory approach, through examples such as the WWW Virtual Library. Indeed, it can be argued that an information source of such size and complexity as the WWW has more to benefit from the application of LIS knowledge than small special collections.

If Ranganathan were alive today he would be aware of the potential of his ideas for searching and organising WWW materials. Indeed, the genius of Ranganathan is attested to by the very portability of his ideas across time, technology and cultures, simply because they address the very foundations of the business of effective information storage and retrieval. Perhaps, in this respect, contemporary WWW developers might find themselves having more in common, or have more time for, Ranganathan, than they appear to have for the current generation of IR researchers, despite their borrowings from the IR research results in the form of the search algorithms embedded in the various search engines. This is largely because contemporary IR research has developed within a particular historical paradigm, the assumptions of which may not fit well with the reality of IR on a widely distributed and disparate information source such as the WWW.

One of the major sources of problems involved in searching the WWW is that in most cases:

> 'there is no closeness at all between designer or creator (which could be anyone) and potential user (which could be anyone or everyone)...the more distant users are, in characteristics and information needs from the types of user conceived of and catered for by those creating or indexing a database, the more likely there are to be problems in accessing relevant information by users from that database'[24].

This is a general problem which is not necessarily resolvable by refining either the indexing or searching on the WWW. Consider the kind of material which would be considered relevant and that which would be considered outlandish in the case of an identical search on the subject of UFOs by a NASA scientist or an X-Files fan – assuming they were not one and the same person!

The problem is of a lack of proximity between searcher and source, of not knowing which item or source will be 'relevant' in a search because of the disparity between searchers and distance between searchers and sources. The problem of indexing for the unknown user destroys or removes the foundations of many contemporary indexing or search algorithms which have to make many such assumptions as the core of their foundations. The fact that the same search would produce completely disparate relevance rankings depending on whether it was undertaken by the NASA scientist or the aficionado of the X-Files is not conducive to the assumptions of contemporary IR research or practice. That is not to say that such techniques do not have their place but the WWW might not be that place.

Facet analysis cannot solve the problem of indexing for the unknown user. However, because it adopts an *a posteori* not *a priori* approach to classification, that is the classification is derived inductively from the concepts or terms used in the subject field, it can be employed to alleviate some of the problems in searching the WWW using either subject directories or search engines. In some respects, facet analysis merges together word and concept based approaches, in that the concept terms in the facets are selected on the basis of literary warrant. The resulting classification should therefore provide a convenient form of organisation of search terms for a subject field. It can also be employed to organise WWW resources, not as a whole, but in relation to organising the results of WWW searches and in relation to web page indexing.

Facet analysis can assist the developer or searcher to retrieval to overcome some of the problems in indexing or searching the WWW in a reasonably effective and efficient way. In other words, it can assist in managing the IR interaction for optimal result, which other IR techniques may not, because of the assumptions made about the IR interaction in their very development. It cannot eliminate the problem of indexing for the unknown user but it can ameliorate it by providing a means of assisting the search process and organising the search results. In terms of searching, facet analysis can be used to suggest, arrange and organise the terms which would be used in a WWW search.

The initial selection of terms for the facet analysis can proceed in a conventional manner by the use of reference works but equally a broad search on the WWW itself is likely to generate a reasonable set of terms for a facet analysis. The terms of the facet classification can be used in the search as well as to organise the search page. The search page can also be made 'live' by the provision of links to sites judged relevant from the search itself. These sites can be linked to the terms of the classification on the screen. This makes the facet display live. In subsequent searches this would enable the searcher to go straight from the chosen search terms to the sites identified or considered as relevant from previous searches. This feature has, in fact, been successfully implemented experimentally in prototype facet search structures by students on the MSc in Information Management at the Faculty of Engineering, University of Porto, Portugal.

Rather than maximising the matching process in order to optimise or refine the identification of 'relevant' documents, which cannot work in the WWW environment in the same way as it might in the test collection environment because of the total lack of control over the creation of WWW material and general lack of proximity between searcher and sources, facet, analysis can be used to optimise the information retrieval interaction. So taking into account both the objective characteristics of the WWW materials and the subjective needs of the searcher. By introducing an element of judgement or knowledge organisation attuned to the searcher as well as to the terms of the subject field the general philosophy underlying facet analysis is more attuned to information retrieval interaction in the WWW environment. Considering those assumptions – (1) Books are for use; (2) Every reader his book; (3) Every book its reader (4) Save the time of the reader; (5) A library is a growing organism – in relation to the problems of searching and organising WWW materials – it is not difficult to see why Ranganathan's ideas should continue to be relevant for IR practice.

## References

1. Ingwersen P. and Wormell, I. 'Ranganathan in the perspective of advanced information retrieval.' *Libri* 42, 1992, pp.184-201.

2. Ranganathan, S. R. *Prolegomena to Library Classification*. (London: The Library Association, 1937.)

3. Ellsworth, J. H. and Ellsworth, M.V. *The new Internet business book*. (New York: J. Wiley and Sons, 1996.)

4. Handley, M and Cowcroft, J. *The World Wide Web beneath the surf*. (UCL Press, 1995.)

5. Anderson, J.D. Indexing systems: extensions of the mind's organising power. In: Ruben, B.D. *Information and behaviour*. (New Brunswick: Transaction Books, 1985, pp.287-8.)

6. Anderson, J.D. Indexing systems: extensions of the mind's organising power. In: Ruben, B.D. *Information and behaviour*. (New Brunswick: Transaction Books, 1985, p.298.)

7. Rowley, J. E. *Organising knowledge: an introduction to information retrieval*. (Brookfield (VE): Ashgate, 1992, p.10.)

8. Shera, J.H. *Libraries and the organisation of knowledge*. (London: Crosby Lockwood and Son, 1965.)

9. Anderson, J.D. Indexing systems: extensions of the mind's organising power. In: Ruben, B.D. *Information and behaviour*. (New Brunswick: Transaction Books, 1985, pp.287-323.)

10. Gilster, P. *Finding it on the Internet: the Internet navigator's guide to search tools and techniques*. (New York: John Wiley and Sons, 1996.)

11. Venditto, G. 'Cybercritic services rate web sites. Which is best? IW Labs rates the raters.' *Internet World*, January 1997. *http://www.internetworld.com/print/monthly/199701/wlabs.html*

12. Tyner, R. *Sink or swim: Internet search tools and techniques* (version 2.11), 1997. *http://www.sci.ouc.bc.ca/libr/connect96/search. htm#search*

13. Sullivan, D. *Search engines features chart*, 1998. *http://searchenginewatch.internet.com./webmasters/features.html*

14. Sullivan, D. *How search engines rank web pages*, 1998. *http://searchenginewatch.internet.com./webmasters/rank.html*

15. Sullivan, D. *How search engines work*, 1998. *http://searchenginewatch.internet.com/webmasters/work.html*

16. Sullivan, D. *The major search engines*, 1998. *http://searchenginewatch.internet.com/facts/major.html*

17. Haskin, D. *The right search engine: IW labs test*, 1997. *http://www.internetworld.com/print/monthly/1997/09/report.html*

18. Binghamton University Libraries. *Comparing search engines*, 1998. *http://library.lib.binghamton.edu/webdocs/search-engine-comparison.html*

19. O'Connor, B. C. *Explorations in indexing and abstracting: pointing, virtue and power*. (Englewood (CO): Libraries Unlimited, 1996.)

20. Buchanan, B. *Theory of library classification*. (London: Clive Bingley, 1979, p.12.)

21. Foskett, A.C. *The subject approach to information*. (London: Library Association, 1996.)

22. Ellis, D., Ford, N., and Furner, J. 'In search of the unknown user: indexing, hypertext and the World Wide Web.' *Journal of Documentation* 54 (1), January 1998, p.28-47.

23. Barlow, L. *The Spider's apprentice: how to use web search engines*, 1997. *http://www.monash.com/spidap4.html*

24. Ellis, D., Ford, N., and Furner, J. 'In search of the unknown user: indexing, hypertext and the World Wide Web.' *Journal of Documentation* 54 (1), January 1998, p.44.

# Send a boy – or dial it yourself? Numbering for the information society

*Ralph Adam*
*City University, Northampton Square, London EC1V 0HB*
*Email: ralph@city.ac.uk*

**Abstract** *Numbering and addressing are key features of telecommunications: without them, the Global Information Society cannot exist. As the demand for new, and more innovative, services grows and the world moves towards deregulation and competition, so the pressure on numbering space increases. This article looks at basic numbering concepts and the role played by numbering in information and telecommunication policy.*

> The descriptions we get of [the telephone's] use in America are a little exaggerated, though there are conditions in America which necessitate the use of such instruments more than here. Few have worked at the telephone more than I have: I have one in my office, but more for show. If I want to send a message, I employ a boy to take it[1].

## The information society

Thus wrote the (British) Post Office's Chief Engineer at what was, perhaps, the start of the information society: a time when library and information systems were well-established; postal, telegraph and telephone services were available, the fax had been invented twenty years earlier and the era of rapid communication of information had begun.

While the arrival of the printing press is often seen as the beginning of the first communications revolution, many present-day commentators consider that the information society will lead to even greater social, economic and political changes.

What constitutes the 'information society', and when it will be fully functional, are moot points: there are almost as many opinions on this as there are commentators. Only four years ago, a report quoted Rupert Murdoch, Wall Street analysts and leading Japanese authorities as agreeing on 2020 as the earliest date for its effective establishment[2].

Most people in the developed world are linked by pairs of twisted copper wires to what, in specialist terminology, is referred to as POTS: plain old telephone services. Yet, despite its unfortunate name, POTS provides important links: conventional voice telephony (with interactive information services), fax, email, classic dial-up online data bases, Internet- and web-access, videotex (e.g. Minitel and Prestel), electronic commerce, interactive video and much more. The problem is that the old-style telephone network is both under-utilised and, yet, too slow for modern communications.

As a result, all societies are confronted by new challenges arising from the convergence of information and communication technology and the growth of multi-media. Competition, liberalisation and innovation, combined with cheaper telecommunication, are seen as the spurs: especially for those applications depending on high-capacity, two-way (broadband) technologies. Digitalisation allows the transmission of both traditional and new services (voice, data or graphics) by varied methods. Some 'old' services using new technology are:

- home banking, home shopping and voice telephony through the Internet;
- email, data and web access through mobile phone networks;
- online services delivered by web-TV or via satellite or cable services[3].

However, in discussing these technologies one must keep in mind the degree to which they are *not* available: over half the world's inhabitants has never made a 'phone call, while two-thirds of telephone lines serve only 14% of the global population, with 63% of countries having no telecommunications outside urban areas[4]. And Manhattan contains more telephone lines than the whole of sub-Saharan Africa[5]. The boundaries of the Global Information Society are the limits to which 'phone lines reach.

## Numbering and the GII

While the elements of the information society are hotly debated, little is said about the 'cement' holding it together. Every communication has a 'sender' (origin) and a 'recipient' (destination), linked by a 'director': rail signal boxes, postal sorting offices, telephone exchanges and Internet protocols all use routing mechanisms to ensure that messages reach their destinations. For electronic communications, a telephone line creates that link: telephone lines need numbers – the forgotten element of the global information society.

A telephone number is a humble, but important, set of digits. It is a prime ingredient of any call: without it, there can be no communication. And a telephone company without numbers can have no customers. Yet, how often do members of the information world think about the origins of numbering and its role in a sophisticated and competitive environment? Very rarely, if the level of debate on numbering policy is any guide. A twenty-year online search of INSPEC produced only seventeen substantive papers: less than an article a year – hardly a case of information overload! Regular conferences take place, newsletters are produced and electronic discussion groups formed, but their content goes mainly to numbering specialists.

This article aims to introduce basic numbering concepts and policy issues to a wider audience, indicating their role within the information society.

Most of the examples relate to the United Kingdom: Britain has one of the oldest, most highly-developed telecommunications systems – notable both for its level of innovation and the competitiveness of its market (currently with well over 150 licensed operators and nearly 600,000 Internet host computers). Also, the British situation will be familiar to many readers. But, wherever you find telephones, there will be 'phone numbers: the issues and problems discussed here are, or soon will be, common almost everywhere.

## What's in a number?

Few people are aware of numbering policy until a change takes place – something which is increasingly common. For example, since 1990, the code for central London has gone from '01' to '071', then to '0171' and, shortly, it will change again to '020-7'.

Conventional telephone numbers have a standard hierarchical structure, with a country code (e.g. '44' for the UK) at the top and as many layers as the national or regional administration deems appropriate. Dialling codes give operators and users information on destinations, routes, service types and costs.

Another category of 'number', which will not be dealt with in detail here, is represented by NSAPs: Network Service Access Points. These can either encompass existing numbering schemes or have a bi-level format, with ISO country codes at the top and nationally-allocated (e.g. BSI) codes underneath. NSAPs can be given to both public operators and commercial organisations of any size[6].

Thirdly, come Internet addresses. There is, generally, a country code (such as '.uk') at the head, although some non-country domains (e.g. '.com' and '.edu') exist and are used mainly by US-based organisations. These, too, have a 'flat' format, the addresses being supplied to Internet Service Providers or commercial organisations. Such Domain Names, intended to make the system user-friendly, translate (through a Domain Name Server) to routing addresses: IP addresses (e.g. '98.37.241.30'). The management agreement for the registration of Internet addresses, hitherto controlled by the US government, was due to expire in September 1998. A new system, allowing for competition in domain name registration, and with a broader range of domains, is under discussion[7].

A problem with Internet addresses (web and email) is that they are unstable (i.e. can disappear without warning) and are not listed in directories. One solution is Universal Resource Numbers (URNs) which can be linked to email addresses, postcodes or other personal numbers.

## Numbers as communication tools

### *The development of numbering*

Telephone numbers were introduced as a standardised shorthand for distant voice communication through a human operator: it saved having to provide the correspondent's name and address. In the early days, no one imagined that large-scale numbering would be required for global communication, nor that vast quantities of telephone numbers would be needed to serve sophisticated computer systems. Now, however, the demand is so great that France Télécom, for example, claims it needs a million extra numbers a year in order to maintain the present level of development[8].

Numbering began on a local basis, with each town or village having its own sequence: often only two or three digits. As telephone density increased, numbers were lengthened with the larger towns and cities going first to four digits and then often being divided into named exchanges with the first two or three letters providing a mnemonic code with local geographical references, such as CENtral, ELYsees or WHItehall.

The introduction of automated subscriber trunk dialling (STD) required codes. The British solution was to take existing place names and abbreviate them into meaningful letter/figure combinations to create a relatively friendly user-network interface. For example, Bristol received BR2 and Brighton BR3, while Leeds and Leicester were LE2 and LE3 respectively. However, most switching systems recognise only numerical information. Therefore, these towns were *actually* treated as 272, 273, 532 and 533 – the digits in the same positions on the dial (the initial '0' in STD codes is an 'escape' from local dialling to other types of call and, therefore, not strictly part of the number).

Soon International Direct Dialling (IDD) became available, producing more complications – partly because there was, at the time, no international standard for the link between letters and numbers on dials and keypads. This, coupled with the exhaustion of mnemonic exchange codes, led most countries to adopt all-figure numbering. In Britain this was done by converting to the numerical equivalents of the letters allocated when STD was first introduced. Hence, the apparently haphazard structure of codes in the UK numbering scheme – for example, (0)835 is St Boswell's and (0)837 Okehampton – 469 miles apart. Furthermore, as new services developed they were given whichever codes happened to be vacant: (0)836 was allocated to the Vodaphone cellular service!

Nowadays, numbering schemes must be able to cope with new and enhanced services, as well as having some logic. One trend (especially in the US) is to bring back alphabetical elements – notably for 'vanity' sequences that make words and cost more to obtain (such as British Midland Airways' UK reservations line: 0345 I-FLY-BM, or Forte Hotels' number: '40-40-40'). Another is for schemes to be 'closed', rather than 'open'. An open scheme has separate dialling procedures for local and toll/trunk calls (with shorter numbers for local, i.e. the most frequently-dialled, calls and an indication of the geographical destination). Closed schemes have a single dialling-procedure for the entire country or region, usually with a standard number length.

Number *length* has become an important issue: the ITU designated the end of 1996 as 'Time-'T', the moment when the maximum digits in the 'international significant number' should be increased from 12 to 15. Operators claimed not to require the extra digits, although Austria, Germany, Hong Kong and Finland were already using them. Expensive changes to, for example, switching equipment, operational support services, and billing systems were needed and many telephone companies have still not bought the necessary equipment, leaving their subscribers unable to dial numbers in the new format[9].

Telephone numbers are also being elongated through the addition of new elements. Carrier-identification codes, are one example, allowing users to select the company through which a call is to be made; end-of-dialling indicators, to assist electronic equipment, are another. As a result, some US subscribers may soon find themselves dialling very long numbers (up to 26 digits in Texas) for international calls. Ironically, while Texans may have the longest numbers, they prefer the personal touch of a human operator[10].

***Where have all the numbers gone?***

Since the eighties, there has been a massive development of telecommunications services, many of them information-related: the rise of 'call centres' and similar services using freephone, shared-cost, premium-rate and DDI lines (Direct Dialling Inwards, allowing staff to by-pass the switchboard) are examples. And this is not to speak of Internet- and web-related services. There has also been a boom in other types of telephone need: fax machines, credit-card validators and modems. Mobile 'phones are everywhere and many households now have different lines for each member of the family. Portable numbers, new entrants to competitive markets and the rapid rise in remote-memory telephone card use have all helped eat up numbering resources. In addition, numbering schemes may be poorly designed and wasteful of numbering space.

***The number stock***

Until recently, it has been a widespread convention that numbering sequences are allocated in blocks of 10,000 – a legacy of the early standard format. First impressions suggest that all the theoretically-available numbers can be used. But, as a rule, operational, technical and psychological factors permit only about 60% of possible numbers to be allocated. There are several reasons for this and, to illustrate the point, Britain can be taken as an example. Here, the format typically consists of the national prefix ('0'), a four-digit STD code beginning with '1' or '2', followed by a six-digit exchange and local number – making ten 'significant' digits in all. Thus, the present code for Brighton is '(0)1273', followed by six digits (larger conurbations have three-digit STD codes with seven-figures for exchange code plus local number). A typical STD area has one million potential numbers ((0)1xxx + six digits). However, the following limitations apply:

1. no number can begin with '0', the 'escape' from local dialling;
2. no number can begin with '1' as this series is reserved for short access codes (e.g. '100' for Operator, '112' for emergency, '1471' for Call-back);
3. local numbers cannot begin with '99' as there is a risk of interference with '999': the pre-1993 emergency number which continues in heavy use, as 'parallel running' with '112', the European standard code, is still allowed;
4. blocks of numbers are reserved for companies wishing to change to DDI, while a separate series (e.g. '3') may be kept aside locally for future lengthening of numbers. Also, changed numbers cannot be re-allocated until a suitable time has elapsed.

Each category accounts for about 10% of the possible numbers, thus giving around 40% of the theoretical numbering stock as unusable[11]. Furthermore, the need to allocate entire blocks of 10,000 to an exchange can lose another 20% of the available stock.

***Broadband services***

The term 'broadband' covers many types of technology, but is generally understood to refer to Asynchronous Transfer Mode (ATM). ATM permits the use of signalling to establish on-demand broadband connections and high bandwidth packet-like switching, with numbering providing the link between the network, users, support and billing systems[6]. Enhanced cable or Internet services require broadband links but the battle with satellite-based services and narrow-band ISDN is far from over.

***IP addresses***

Internet telephony is increasingly viable and work on standards well developed. However, if calls are to be made between traditional telephone subscribers and Internet users, it will be necessary to use conventional numbers, as IP addresses cannot be dialled. Project TIPHON, a study by ETSI (the European Telecommunications Standards Institute) is investigating ways of overcoming this[12].

## Non-geographic numbers

### *Information and other special services*

The use of telecommunications as a (profitable) business and information tool has, in recent years, led to the creation of many types of non-geographic number: the code indicating the *type* of call, but not its *destination* – the initial digits showing both network and caller who pays and at which rate. Such services tend to be Intelligent Network-based. For example, a local-rate call from London, using British Airways' reservations line will be answered by a clerk in Scotland during the day or by one in New York at night. From the enquirer's point of view, the link is transparent and only the clerk's accent tells him whether he is through to London, Glasgow or New York.

The main types of number used for special services are:

*Freephone*
Here, the called party pays the entire cost. A recent creation has been the Universal International Freephone Number which, in principle, allows a caller to use the same number from any country in order to access a service. The debates on setting the UIFN standard were heated: delegates from North America, where the majority of freephone numbers are, wanted to use their existing numbers, prefixed by the Country Code. Representatives from the rest of the world fought this on the grounds that it would give the USA a clear competitive advantage[12].

*Shared-cost*
Callers pay only part of the conveyance charge, with the remainder, and any extra for the use of Intelligent Networks, being covered by the called party. The cost of a call is the same, irrespective of point of origin. Such services are often charged at 'local' or 'national' rate.

*Fixed-cost*
These allow organisations, such as national information services, to use Intelligent Networking to send calls to wherever staff are available. The enquirer pays a fixed rate irrespective of actual destination, with the called party paying the balance[13].

*Premium rate (shared-revenue)*
Such numbers indicate that the cost of the call will be above a certain level, with revenue shared between service provider and user. Telephone companies have traditionally been much more concerned with how, rather than what, they are carrying. Now, however, content has come into question and premium-rate lines, which often carry recorded messages, have increasingly come under the spotlight for dubious content[14] (as, too, have online services[15]).

Standards for both international shared-cost and international shared-revenue numbers are currently being finalised by the ITU, although content standards continue to be a national or regional matter.

*Personal communications*
Mobile telephone and paging services, for which several competing standards exist, have developed dramatically over the last decade – especially in Asia and northern Europe. Most authorities treat them as non-geographic, with charging systems that are not distance-related, and each operator having separate number ranges. Numbering problems arise when 'phones are taken abroad – which country's code is to be used? An international code for such purposes might prove a popular option.

*Number mobility*
Personal ('follow me') numbering. This allows people to 'buy' numbers for life which can be routed to any phone (for example, home, library or holiday cottage). Such numbers will increasingly have added services, such as voice mail.

***Portability***
Changing numbers when moving house or telephone company is inconvenient and costly (stationery must be reprinted, change-of-address information sent out, fax machines, modems and alarms re-programmed and calls transferred). *Number portability* refers to the possibility of taking a number on a long-term basis from one place, or one operator, to another, with significant savings and convenience resulting. Such services are a key feature of competition and can only operate where there is a neutral administrator or regulator. So far, three main forms have been introduced: *provider* portability (for changing telephone companies), *service* portability (the transfer of 'enhanced' facilities) and *location* portability (when moving geographically).

The idea of keeping one's number when moving sounds attractive but, in practice, has resulted in many political and organisational issues – inter-company conflicts, disagreements with regulators, accusations of anti-competitive behaviour and pricing problems. There are also disagreements as to whether mobile and non-geographic numbers will, eventually, become fully portable[16,17].

## Controlling the numbers

### *The role of the Regulator*

The old approaches to national regulation, favouring stability and support for domestic industries to the exclusion of foreign competition, are fast disappearing. In most countries, pre-deregulation, number allocation was carried out by the national operator or the relevant Ministry (the ITU performs this role for the global scheme); in deregulated environments independent, impartial, bodies must take a wide view, ensuring conflicts of interest (of which there are many) are resolved and that well-designed, stable numbering schemes, with effective consultative mechanisms, are created (within the EU, national regulators must comply with the Council's policy framework and the decisions of EC-recognised bodies). Thus, the Regulator should:

1. maintain a long-term view of the numbering scheme, foresee capacity shortages, review needs and select the scheme's architecture;
2. consult all interested parties regularly and, in these consultations, reflect users' interests;
3. decide on, and publicise, the rules governing use of the numbering scheme: number structures and lengths, ranges for use or holding back, allocation to operators, harmonisation, competition issues and so on;
4. administer the scheme: allocating number blocks to network operators and withdrawing unused or badly-used blocks;
5. set rules relating to other relevant competition issues, such as equality of access and number portability;
6. resolve disputes between operators – over number allocations, for example[18,19].

In contrast with traditional telephony, the IP environment is unregulated. Telephony over the Internet raises new issues for regulators. Companies providing the service will be competing against conventional telephone operators, some of which will be offering it themselves. Commercialisation of Domain Name control requires a major revamp of DNS regulation[20].

## Standards

Much has already been said about standards. Current attitudes recognise that telephone numbers are a scarce national resource which can neither be 'owned' by individuals nor by telephone companies (a major reason for the separation of service providers and numbering authorities in deregulated economies). The supply of numbers is finite – they must be conserved with safeguards for future capacity demands and new services, the need for which may not yet be apparent. Furthermore, plans for the future must support effective competition by ensuring that access to numbering resources is equal for all service suppliers.

Standards, always a key element of telecommunications policy, have become an increasingly important issue over recent years. International standards are crucial for network interconnection and manufacturers have battled to have their own products accepted as *de facto* standards. However, in numbering, the official bodies still have a key part to play.

The International Telecommunication Union (ITU), is the UN specialised agency covering all aspects of telecommunications. Since its 1993 re-structuration, the organisation has had its own Telecommunication Standardisation Sector, known as ITU-T. This is divided into 15 Study Groups, each handling specific areas (Study Group 2, responsible for Network Operation, covers numbering issues), with administrative activities carried out by the Telecommunication Standardisation Bureau (TSB). These activities are described in detail by Channing[21].

Although the make-up of ITU-T is such that a wide range of views (especially those of telephone operators, manufacturers, users and regulators) is represented, it is the commercial element that acts as the driving force[12]. There is little input from the information world.

Every telecommunications operator has an obligation to route calls appropriately. The international numbering plan, contained in ITU Recommendation E.164, provides Country Codes (CCs) of from one to three digits, in front of national numbers. The emergence of new states, and the consequent need for more country codes, has raised the possibility of recalling all one-, and even two-, digit numbers. The present arrangement allows for only about 80 more CCs to be created (the break-up of the Soviet Union and Yugoslavia depleted the stock). When a Country Code changes, everyone has to adapt and the cost of informing the rest of the world, reprogramming switches and adapting information systems is high.

Deregulation and liberalisation have brought important changes to standards, many initiated by manufacturer-driven industry fora 'feeding in' decisions to international and regional bodies, such as the ITU and ETSI. Conflicts exist between standardisation bodies: for example, the ATM Forum (which bases its proposals on the ISO's NSAP addressing structure) and the ITU (with a preference for its own numbering scheme) have both made recommendations on ATM addressing[6]. The two groups are now trying to resolve their differences[12].

The present generation of standards is seen as a foundation for a market, rather than a technology, driven future[22].

## Harmonisation

This is a much-discussed topic and proposals include standardising trunk and international access codes, short codes and non-geographic services. The obvious questions are: who do you harmonise *with*? And who benefits?

The European Union has been trying, with only limited success, to harmonise telephone numbers. Large countries can have numbers of up to eleven-digits and it is hard to make these memorable. Short codes of two or three digits are much easier to recall and tend to be used for information and emergency services, but they differ widely between countries. Current policy is to standardise them (so that, for example, '112', rather than '13', '15', '17', '115', '998' or '999', calls the emergency services). Such numbers are, however, very wasteful of numbering space: up to 100 million numbers can be 'lost' behind a single short code. A pan-European Country Code, similar to '1' for North America, has also been proposed, but has been shelved due to opposition from both the industry and users.

New switching and transmission technology have encouraged the simplification of network structures and reduced the number of exchanges, administrative areas, charging and numbering zones. The relationship between income from geographic and non-geographic calls is also changing, especially in the USA, where more than half of all trunk calls are freephone[23]: there is, consequently, a shortage of freephone numbers, with operators being accused of hoarding them.

## And....what about the users?

Virtually everyone in the developed world uses telephone numbers and IP addresses, yet user studies are almost unknown in the numbering policy arena. Users' needs have rarely been discussed and few professional papers have addressed the issue[24]. What *is* known of users' preferences suggests that numbers should be:

1. *dependable* – easily retrieved or remembered, reproduced accurately and infrequently changed;
2. *meaningful* – standard lengths and uniform patterns and formats (e.g. 3+3+4 or 3+4+4 digits) are preferred, even if it means dialling extra digits: the benefits are increased user confidence and fewer misdialled numbers. Long numbers are impersonal and not user friendly: people cannot identify with them as local features. They are also less memorable (psychologists claim seven to be the optimum number of digits for the average person to recall[25]);
3. *correctly received* – numbers should not be readily mis-dialled from other much-called numbers;
4. *consistent* – the same for long periods (even if the user moves and/or changes operator);
5. *recognisable* – with easily identified codes and giving appropriate information (e.g. likely call cost or geographical location).

Users are not agreed on everything. Differences of opinion can be summarised as:

1. having the numbers as short as possible seems to matter much more to residential than to business users, or to those in countries with long trunk dialling codes. Business users often favour standardised dialling, irrespective of destination;

2. the significance of special numbers varies according to cultural factors: for example, many people in Asia consider certain numbers to be lucky or unlucky;
3. views on international harmonisation and standardisation vary according to how much individual users travel abroad or the extent to which they have an international outlook on life[18].

Future developments should satisfy some, at least, of these preferences.

One possible reason for lack of concern with users' needs is that professional groups (administrators, engineers, marketers and policy-makers), themselves, feel misunderstood by one another[17]!

## Numbering *and* the information society

The aim here has been to give an aerial view of numbering: too broad a topic to be encompassed in a single article. Key aspects can be followed through from the references. Because numbering affects so much of life in developed countries, the subject has been presented in general terms, without specifying in detail issues relevant to library and information work. Many such links are obvious.

There are, however, several areas which merit further attention. One is the role of information professionals, key workers in the information society, in policy-making. Another is the degree to which debate has, so far, been limited to transport mechanisms, rather than the data and information being transported. This is rather like discussing library buildings, without considering the stock held within them.

### *Librarians in numberland*

The information world has much to contribute to the, as yet, undeveloped area of user studies. Numbering schemes are changed, global numbers created and domain names restructured, but there is little understanding of what the GII is or what needs to be done to fulfil the original vision. While it is claimed that naming and addressing work should result in a common understanding of the evolutionary path to be followed in order to meet these needs[12], where is the research on what *users* need and want?

Also familiar to the information world are those issues concerning intellectual property (conflicts between trade-mark holders and 'owners' of disputed domain names or vanity telephone numbers) and privacy (with respect to calling-line identification, for example).

Documentation is another issue: how can 'phone books best be organised when numbers begin with carrier identification codes which are, in effect, advertisements for telephone companies[26]? Or, how should information on number changes be conveyed?

There are also areas from which information professionals can benefit. Intelligent networks, allowing calls to be handled virtually anywhere from a single number, could allow inter-library collaboration in round-the-clock enquiry services. The call-centre movement (again, an unresearched topic) uses such services extensively. In Britain there are at least half-a-million employed in call centres which, increasingly, use information scientists to raise quality: one insurance company, PPP, uses medical librarians in its call centres[27].

Telecommunications have come a long way since 1879, yet much remains to be done in the development of numbering policy. But, with the USA running out of numbers fast, and Texans building longer and longer numbers, perhaps the descriptions we get of the telephone's use in America are not exaggerated after all!

## References

1. Pearce, W. *House of Commons Special Committee on the Telephone*. (Submission by Chief Post Office Engineer. London: HMSO, 1879.)

2. Federal Trust. *Network Europe and the Information Society*. (London: Federal Trust For Education & Research, 1995).

3. CEC. *Green paper on the convergence of the telecommunications, media and information technology sectors, and the implications for regulation: towards an information-society approach.* (Brussels: Commission of the European Communities, 1997).

4. Graubner, J. 'In search of the dial tone.' *Asian Communications* 12 (8), 1998, pp.46-50.

5. Schlögl, C. 'A review of the information and telecommunications policy of the European Union.' *Review of Information Science* 2 (1), 1997, pp.1-5.

6. Taaffe, K. 'Broadband numbering and addressing.' In *Proceedings of Numbering '98.* Edited by N. McHugh. (London: Phillips Omnicom), 1998.

7. NTIA. *Management of Internet names and addresses*. (Washington, DC: US National Telecommunications and Information Administration, Dept of Commerce, 1998.)

8. Adam, R. 'The new French numbering plan.' *Telecommunications Numbering Review* 1 (7), 1996, pp.4-7.

9. Cuccia, M. *Longer-than-twelve-digit countries?* Posted on: *telnum-l@else.net*, 24th July 1998.

10. Connors, R. 'Numbering issues for today and tomorrow.' In *Proceedings of Numbering '96.* Edited by M. Corby. (London: CommEd, 1996.)

11. Oftel. 'History of the UK numbering scheme.' In *Geographic telephone numbers – a consultative document*. (London: Office of Telecommunications, 1995.)

12. Holmes, T. 'International numbering issues – an introduction.' In *Proceedings of Numbering '98.* Edited by N. McHugh. (London: Phillips Omnicom, 1998.)

13. Orbell, A. 'Numbering' In *Telecommunications Users' Handbook*, eighth edition. Edited by M. Corby. (London: Phillips Tarifica, 1998, pp.6.2-6.34.)

14. Nickelson, R. 'The global challenge – strategies in transition.' In *Proceedings of Telecom '95: the Seventh World Telecommunication Forum.* Strategies Summit (Supplement): Session 22, paper 5. (Geneva: International Telecommunication Union, 1995.)

15. ABA. *Investigation into the content of on-line services: report to the Minister for Communications and the Arts.* (Sydney: Australian Broadcasting Authority, 1996.)

16. Reinke, T.H. 'Local number portability and local loop competition: critical issues.' *Telecommunications Policy* 22, 1998, pp.73-8.

17. Adam, R. 'Number portability in the Asia-Pacific Region.' In *Proceedings of Numbering '98.* Edited by N. McHugh. (London: Phillips Omnicom, 1998.)

18. Milne, C. 'The design and management of numbering systems.' In *Telecom reform principles, policies and regulatory practices.* Edited by W. H. Melody. (Lyngby: Technical University of Denmark, 1997), Chapter 12, pp.149-63.

19. Ha, Y-K. 'Development of a numbering policy for Hong Kong's competitive telecommunications environment.' *Telecommunications Numbering Review* 1 (8), 1996, pp.3-10.

20. Massonet, E. 'Naming and addressing for the Internet: a regulatory issue?' In *Proceedings of Numbering '97: the Seventh International Numbering Conference*. Edited by M. Corby. (London: CommEd, 1997).

21. Channing, I. 'Making the connection: the ITU's standardisation activities.' In *International Tele-communication Union: Celebrating 130 years.* Edited by I. Channing. (London: International Systems and Communications, 1995), pp.140-153.

22. Sharrock, S. 'Evolving standards.' In *International Telecommunication Union: Celebrating 130 years.* Edited by I. Channing. (London: International Systems and Communications, 1995), pp.418-421.

23. Adam, R. Freephone goes invisible. Icbtollfree, 24th July, 1998. (*http://www.icbtollfree.com*).

24. Adam, R. 'Numbering goes European.' *Telecommunications Numbering Review* 3 (4), 1998, pp.3-7.

25. Miller, G.A. 'The magical number seven, plus or minus two: some limits to our capacity for processing information.' *Psychological Review* 63, 1956, pp. 81-97.

26. Cox, R. 'Le renumbering – a commentary on the French plans.' *Telecommunications Numbering Review* 1 (8), 1996, pp.12-15.

27. Cobb, R. 'Telephone extension.' *Marketing*, 10 September 1998, pp.38-42.

# Football information services: fanzines, 'Match of the Day' and the modem

*Richard Green*
*Disclosure First Contact, 1 Mark Square, London EC2A 4PR*
Email: rgreen@disclosure.co.uk

**Abstract** *The character, growth and demand for football information services in England and Wales is examined. Questionnaires were sent to football clubs and supporters to investigate the quantity, quality and types of information services available, and whether the demand for football information is being met. It was found that there is more football information available than previously, through a wider range of services, and that demand for club information is growing. However, supporters feel the information is often of poor quality. The growth in football information services is attributed to the appeal of the game to non-traditional supporters and the use of information services to market clubs as global 'brands'. Democratic supporters movements have had a major influence on the development of football information services. The Internet is set to have a big impact.*

## Introduction

According to a recent report, 18 million people consider themselves to be football fans, more than a third of the population of England and Wales[1]. All the evidence suggests the game is experiencing unrivalled levels of popularity. The increase in interest has coincided with technological and regulatory changes making possible new and radically different forms of communication, both directly between football club and supporter, and across a wide sector of the media in general. Football clubs are more dependent than ever on information services to market their particular 'brand' in order to attract television coverage, merchandising sales and corporate hospitality, as well as to attract new supporters and communicate with existing ones. Yet despite this football information services have obtained virtually no attention from researchers. Most academic research on football has concentrated on sociological aspects (violence, crowd behaviour, social history), or physiological aspects (specifically sports injuries). Other research has, to a lesser extent, investigated the connections between football and popular or youth culture, particularly the influence of music and fashion. With the recent introduction of new forms of information delivery technology – particularly those associated with the Internet, including the World Wide Web, discussion groups, live commentary relay over the Internet, and satellite television, which offers the possibility of football clubs controlling content on their own television channels, the need for research is all the more pressing.

With this in mind an investigation was undertaken:

1. to map and assess the information services available, with regard to their range, type and quality;
2. to identify whether the need for football information is being met, and if it is likely to grow in the future;
3. to identify the factors influencing demand for football information services and in particular the commercial forces behind such demand;
4. to examine the role of football supporters in influencing the provision of football information services.

The study was limited to information services provided by or about professional English League association football clubs. Leading media figures, supporters representatives, football authorities, information service personnel and football club officials were all questioned.

## Methodology

Questionnaires were mailed to all English and Welsh League clubs and delivered by hand to a small sample of football supporters. The club questionnaire was sent to all 92 English and Welsh league clubs. This questionnaire sought to determine which information services were provided by football clubs. Further questions sought to examine the attitudes of football clubs towards official and unofficial sources of football information, and to external information sources (print media, television, fanzines, football magazines). The clubs were also asked to assess their services and comment on future plans. Fifty-two completed questionnaires were received, a 57% response rate. This response rate was consistent throughout the four divisions with the exception of Division 1, which was significantly lower: Premier League 12 out of 20 clubs (60%) returned questionnaires, Division 1 8/24 (33%), Division 2 16/24 (67%), and Division 3 16/24 (67%).

The questionnaire aimed at supporters was distributed in London and Manchester. It was felt that this approach avoided potential bias attached to one region. Seventy-two questionnaires were distributed and 28 returned – a 39% response rate. Information was sought about which services fans used; which services they thought their club provided; the type of information they sought; how often they used the services; how much they paid and were willing to pay; quality of services; accuracy, currency and value for money of services; growth in quantity of services; official and unofficial information; and opportunities for feedback.

Interviews with representative figures from football and the media were conducted in order to provide a context for the questionnaire data, and to examine football information services from sources other than football clubs. Interviewees were chosen to reflect all aspects of football information provision and demand, with an emphasis on those directly involved in the process or possessing acknowledged expertise. These included journalists, broadcasters, freelance writers, football club officials, museum and archive curators, commercial service providers (websites and phone lines), academics, football administrators, supporter association officials and individual fans.

## Football information services

### *Non-club services*

#### *Fanzines*

Fanzines are generally low budget limited distribution magazines or pamphlets produced by individuals, or small groups of enthusiasts, on the subject of their particular hobby or pastime. Perhaps their most important characteristic is their unofficial and independent status. According to Mark Wylie[2], Manchester United museum curator, a fanzine is simply 'an outlet for people to have their say'. The first football fanzines appeared in the mid 1980s, early examples being Bradford City's *City Gent* and the general football fanzine *When Saturday Comes*[3]. In terms of style and production values they had antecedents in the alternative underground press of the 1960s, and in particular the numerous fanzines that sprang from the punk music era of the late 1970s[4,5]. Of particular relevance to the later football fanzines was the way in which these tended to democratise a particular lifestyle and passion [6,7]. Most fanzines are locally produced and linked to a particular club[8].

The fanzine archive at the Manchester Institute for Popular Culture contains thousands of titles, many short-lived, some still publishing after more than ten years. Today few, if any, clubs are without a fanzine, and most have several alternatives. As an example, the Institute holds copies of nine different titles for Manchester United. *United We Stand* sells 8,000 copies each issue; *When Saturday Comes* is now distributed nationally with a readership of over 40,000. For many fanzines, distribution is by hand outside the ground on match days; however, Haynes[6] has pointed to a distinctive distribution network centred around specialist book shops and independent record shops.

By adopting an argumentative and critical unofficial discourse fanzines have often found themselves in opposition to the clubs they support. Nonetheless, they have had an enormous influence on the way the game is represented in the media and official discourse. Firstly, they set the tone for a new style of football writing, in which the fans' point of view is paramount. Secondly, fanzines have been a major force behind the recent trend towards greater democracy and participation in footballing discourse, as part of a concerted effort to reclaim the game for the ordinary supporter[9]. Partly in response to the negative media image and public reaction following disasters at Heysel and Hillsborough, and reflecting the rise of the football supporter associations, fanzines have encouraged dialogue and an expression of views unmediated by unsympathetic journalists, governing bodies, the official club line or the needs of sponsors. Above all else, fanzines are documents providing an alternative viewpoint.

### *Television*

Until comparatively recently the football authorities were reluctant to sanction live coverage of games, fearing loss of attendance and club gate revenue. The poor image of football did little in return to attract broadcasters and advertisers. Two indirectly linked events in 1992 provided a major boost to growth in demand for football: the formation of the Premier League, and the awarding of live television rights to BSkyB. Sky desperately needed a strong attraction to boost subscription sales, and football provided it. According to a 1997 report from Case Associates, 'Football is crucial to pay tv, and BSkyB in particular. It is one of the 'drivers' or 'killer applications' which make people want to subscribe'[10]. Ironically Sky executives had originally thought that films were going to be the major drivers. Subscriptions jumped from 1.8 million in 1992 to 6.4 million in 1997[11].

The success of live television coverage depended largely on its ability to deliver two crucial audiences. For advertisers this was the young, white, middle class male with disposable income. For broadcasters televised football attracted the 16-24 year olds who might not otherwise watch much television[12]. The result has been a commercial success which has in turn widened the market for clubs to exploit through their merchandising operations. In 1997 Sky more than doubled their bid to £670 million. Media consultancy Oliver & Ohlbaum have suggested that the deal is actually worth £1.3 billion to Sky over the same period[13].

Many supporters have expressed fears that there is now too much football on television[14], and that the influence of television has been largely negative. Nonetheless, fears that televising live matches would lead to a fall in attendances have proved groundless and live coverage has helped generate demand for the sport as well as reflecting existing demand.

### *Newspapers*

The initial growth in popularity of professional football at the end of the last century took place at a time when newspapers were still the prime source of mass information dissemination. Specialist local sportspapers such as *The Athletic News* or *The Sporting Chronicle,* aimed at the male working class supporter, sprang up to meet the nascent demand[15]. It was not until the 1920s and 1930s that sport, and football in particular, were established on the back pages of national newspapers[6]. Newspapers occupied a near monopoly position in the provision of football information which they guarded closely, expressing opposition to early experiments in both radio and television sports broadcasts[16].

Following England's World Cup victory in 1966, football achieved levels of popularity greater than at any time since the immediate post WW2 peaks and, in the process, cemented its place as the national sport. The effects of the abolition of the minimum wage for players in 1961, in concert with wider societal changes (especially the emergence of popular youth culture), created the first football 'superstars'. Newspapers began to reflect the general interest in players' lifestyles[15]. Off the pitch fans became the focus of football reporting for the first time as violence between spectators spread at football matches in the late 1960s and early 1970s. Resentment by the majority of law abiding supporters at this negative portrayal in the media led ultimately to two democratic movements which aimed to give a voice to genuine supporters which had been excluded from newspapers: independent supporters associations and football fanzines. The fanzine style made little immediate impact, but after the great success of Nick Hornby's *Fever Pitch*[17] editors were quick to seize on the possibilities and popularity of this new approach to football writing, recruiting a new generation of 'Post-Nick Hornby' football writers[14]. This shift in emphasis was also a belated recognition that newspapers had been overtaken as the first source for hard and current news; fans are now informed by a multiplicity of media, and want value added reports from newspapers rather than the scores they already know[18].

The change in style and scope of football writing has been matched by an increase in volume. Commercial success elsewhere – sales of *Fever Pitch* and fanzines, Sky tv subscriptions, club merchandise – proved that football sells. Newspapers have responded by expanding the back pages coverage, recruiting more, younger writers, and introducing dedicated sports sections. Football coverage also appears more often outside its traditional place on the back pages.

There is, therefore, not only more football coverage in newspapers, but various levels of information aimed at different sections of the population. The style of much newspaper coverage has changed, reflecting the fanzine's mission to represent the fan's viewpoint and reveal what goes on behind the scenes at football clubs. The accuracy and quality of such information is open to doubt, especially in the tabloid press. A number of pressures contribute to the perceived level of misinformation. Firstly, there is the pressure of finding enough accurate information to fill the necessary column inches. Secondly, information has more filters to pass through; clubs, owners, shareholders, sponsors, players and players' agents all have a vested interest in manipulating information in their own interests. Thirdly, clubs have become more adept at PR and in handling the media. Share prices, ticket sales and merchandise sales can all be affected by the selective leaking of information and misinformation. A disturbing recent trend is for clubs to ban reporters who have written negative articles, leading to self-censorship by journalists whose contacts or even livelihood are at stake[14,18].

### *Those provided by the clubs*

Ten different types of information service are provided by the clubs (*Table 1*). All 52 responding clubs offered match programmes. Next most common was the premium rate phone line, offered by 49 of the clubs, including all Premier League, Division 1 and 2 clubs. Third most common was the merchandise catalogue (38 clubs), though this ranged from over 90% of Premier League clubs to only 50% of Division 3 clubs. The least common services were radio stations (9 clubs) and yearbooks (14 clubs).

*Table 1. Information services offered by clubs*[1]

| Service | Total | Prem | Div1 | Div2 | Div3 |
|---|---|---|---|---|---|
| *Number of clubs* | *(52)* | *(12)* | *(8)* | *(16)* | *(16)* |
| Match programme | 100 | 100 | 100 | 100 | 100 |
| Premium rate phone line | 94 | 100 | 100 | 100 | 81 |
| Merchandise catalogue | 73 | 92 | 88 | 75 | 50 |
| Fan forums | 69 | 75 | 88 | 56 | 69 |
| Ticket/Match phone line | 64 | 92 | 88 | 63 | 31 |
| Information brochure | 54 | 83 | 75 | 44 | 31 |
| Official website[2] | 50 | 83 | 63 | 38 | 31 |
| Official magazine | 39 | 75 | 63 | 25 | 13 |
| Yearbook | 27 | 58 | 38 | 6 | 19 |
| Club radio station | 17 | 33 | 13 | 6 | 19 |

1. Expressed as a percentage.
2. As of 8/8/98 all bar one club has a website.

On the whole clubs felt that they did a good job in providing information about the club to supporters, although a significant proportion thought their performance was just average (*Table 2*).

*Table 2. Club satisfaction with their information services*[1]

| Rating | Total | Prem | Div 1 | Div 2 | Div 3 |
|---|---|---|---|---|---|
| Excellent | 14 | 18 | 25 | 13 | 6 |
| Good | 63 | 82 | 63 | 56 | 56 |
| Average | 24 | 0 | 13 | 31 | 38 |
| Poor | 0 | 0 | 0 | 0 | 0 |
| Very poor | 0 | 0 | 0 | 0 | 0 |

1. Expressed as a percentage.

This general sense of satisfaction no doubt contributed to an optimism about the future: 72% of clubs anticipated generating more revenue from information services in the future; none anticipated generating less.

P9474

Two of the information services – premium rate phone lines and the Internet – are relatively new and novel and deserve special attention.

*Premium rate phone lines*

One of the most widely available and directly accessible of the commercial football information services is the premium rate phone line. For many years a limited range of services was available on the telephone, such as the speaking clock, weather reports and dial-a-disk. Legislative changes in the early 1980s opened up the possibility of premium rate lines. The first sports service was a phone betting line operated by the bookmakers William Hill, who went on to introduce live commentary on horse races[19]. In the mid 1980s BT introduced the first football lines known as ClubCall. Initial reaction from football clubs was hostile; many clubs could not understand how they could make money and argued that their supporters would not want to pay to make phone calls to find out information about their clubs. They were proved wrong. By 1993 ClubCall (which had been bought by Ladbrokes for £40 million) was paying royalties of £7 million per season to football clubs[20]. Not surprisingly then, 98% of the clubs questioned said they operated premium rate phone lines.

For many lower division clubs who lack national media coverage the phone lines provide a ready source of both revenue and information dissemination, especially for supporters living outside the local area. The lines mainly provide: up to date news stories (98% of respondents); interviews with players and managers (92%); upcoming fixture details (92%); and match reports (86%); they may also contain general gossip, particularly relating to transfer news. Least commonly provided was material relating to club history and archives (12%) and the opportunity for supporter feedback (12%). Plainly it is currency that is being sold over the telephone.

*Websites*

The Internet provides football fans with cheap, instant and global communication opportunities, enabling supporters to exchange information and ideas. Use of the Internet has been documented both in arranging friendly meetings between fans from different countries prior to European competition matches[21], and as a means of arranging fights between rival supporters[22,23]. In recent seasons the Internet has been used extensively to coordinate protest campaigns by supporters of Brighton and Doncaster Rovers, extending the campaigns to football supporters nationwide. Club websites serve supporters with a constant source of information on their team even when living outside the local area, and especially for the exiled fan abroad. More than half the visits to Teamtalk's website are from overseas; this can help to sustain the relationship between club and supporter as well as provide direct marketing opportunities[19].

The democratic nature of communication and information exchange is also matched by the non-hierarchical structure of the web. Hypertext links allow fans to navigate their own paths through the masses of sites devoted to football, allowing for a flexibility and multiplicity of choice that suits the voracious, and often haphazard, information seeking behaviour of football fans. Despite the thousands of football websites on offer, football clubs themselves have been slow to take advantage of the commercial opportunities offered by the Internet. Many clubs, especially outside the Premier League, still have no official website. Just 45% of the clubs questioned claimed to have an official website. Broken down by division, 75% of Premier League clubs have official websites, 89% of Division 1 clubs, 50% of Division 2 clubs and 27% of Division 3 clubs. However, the situation is fast changing with half of those without one planning to add a website. Websites offer commercial opportunities for clubs in three distinct ways. Firstly, they can be used as a vehicle for selling advertising or 'banner' space. The use of the Internet as an advertising medium has been slow to date, across all fields, but is widely expected to improve. Secondly, websites can be used to advertise and sell club merchandise directly – one of the fastest growing areas of club revenue in football. Thirdly, the Internet provides the means to communicate and market brand identity and build loyalty on a global scale, creating awareness of both individual clubs and English football in general.

Websites allow clubs – from the Premier League to the lower divisions – direct control of content and information rather than having to rely on newspapers and other external media.

As clubs seek to increase control over the content of information provision, and exploit its potential for communication with supporters, football information on the Internet is likely to increase. Technical developments already underway will add to its capabilities. The immediate future will see more clubs establishing official websites with the aid of professional web page designers and providers. These sites are likely to match the best designed unofficial supporters' sites, provide more multimedia information in the form of audio and video clips, and offer an extended range of services including merchandise and ticket sales. The official website of the future is likely to feature virtual reality stadiums where supporters can test the view of the pitch from different seats before purchasing them[24].

Websites are also likely to provide access to the ultimate information product controlled by football clubs, the match itself. Chelsea fans world-wide can listen to live radio commentary on their team's matches through the Internet[25,26]. A number of other clubs have taken their lead. Technical improvements have made live video possible, and despite contractual problems over the rights to live images, several clubs (notably the Caspian group which owns Leeds United) have expressed interest in screening matches live on the Internet[27].

One major obstacle to be overcome if the Internet is to become a mass medium for football information is that of access. At the moment just 12% of British homes have a computer. The development of set-top boxes providing Internet access through televisions will revolutionise the commercial appeal of websites, and the first such boxes are already appearing on the market[28]. Other technical developments in the security of online financial transactions, and new charging mechanisms – charging and collecting fractions of a penny for small pieces of information, such as a web page or player photo – look set to improve the commercial viability of the Internet for football information.

## Information needs

### *Quantity of information*

The vast majority of fans (89%) felt that there was more football information around today than ever before (*Table 3*). Despite the greater volume, just under a third (30%) of fans felt there was not enough information available about their club (*Table 4*). The information gap was felt more strongly in the lower divisions (60% of fans of clubs in Divisions 2 and 3 felt there was insufficient information).

*Table 3. Volume of football information compared to previous years*[1]

| Quantity | Total | Prem/Div 1 | Divs 2/3 |
|---|---|---|---|
| More | 89 | 89 | 90 |
| Same | 11 | 11 | 10 |
| Less | 0 | 0 | 0 |

[1] Expressed as a percentage.

*Table 4. Amount of information available about the club*[1]

| Quantity | Total | Prem/Div 1 | Divs 2/3 |
|---|---|---|---|
| Not enough | 30 | 12 | 60 |
| Sufficient | 67 | 82 | 40 |
| Too much | 4 | 6 | 0 |

[1] Expressed as a percentage.

Where some fans thought there was too much information was in regard to football in general. But they were in the minority (16%) and almost matched by those who thought there was not enough (12%). And the position was more polarised in the lower divisions where only half the fans felt there was sufficient information and a quarter each felt there was either not enough or too much.

*Table 5. Availability of information about football in general*[1]

| Quantity | Total | Prem/Div 1 | Divs 2/3 |
|---|---|---|---|
| Not enough | 12 | 6 | 25 |
| Sufficient | 72 | 82 | 50 |
| Too much | 16 | 12 | 25 |

[1] Expressed as a percentage.

### *Type of information*

Fans largely sought news (100%), results (surprisingly only 92% of fans), fixtures, match reports and transfer news (85% each). It was primarily current and fast changing information that was in high demand. The major difference between divisions was that fans from the lowest divisions were totally uninterested in club history – perhaps because their clubs were less glorious. Plainly fans from all divisions do not view the merchandising catalogues as anything other than a marketing tool and had little inclination to seek this kind of information out.

*Table 6. Type of football information sought*[1]

| Type of information | Total | Prem/Div 1 | Divs 2/3 |
|---|---|---|---|
| News | 100 | 100 | 100 |
| Results | 92 | 94 | 89 |
| Fixtures | 85 | 82 | 89 |
| Match reports | 85 | 82 | 89 |
| Transfers | 85 | 76 | 100 |
| Statistics | 46 | 53 | 33 |
| Interviews | 46 | 41 | 56 |
| Analysis | 46 | 53 | 33 |
| Supporter feedback | 38 | 35 | 44 |
| Club history | 15 | 24 | 0 |
| General history | 12 | 18 | 0 |
| Chat/interactive | 12 | 12 | 11 |
| Merchandise info. | 4 | 6 | 0 |

1. Expressed as a percentage.

### *Authority*

The fan's general preference was for information from official sources, though less than half (46%) of the fans said so (*Table 7*). In fact the lower division fans preferred unofficial sources (40%) to official sources (30%). Perhaps the relatively low showing of official information is due to a lack of authority and honesty on the part of club officialdom. A common complaint of fans was that information was often of poor quality and not always accurate.

*Table 7. Preference for information from official or unofficial sources*[1]

| Source | Total | Prem/Div 1 | Divs 2/3 |
|---|---|---|---|
| Official | 46 | 56 | 30 |
| Unofficial | 29 | 22 | 40 |
| Both | 25 | 22 | 30 |

1. Expressed as a percentage.

### Information use

Overall the most used information services for football were match programmes (89%), national and local papers (both 75%), Teletext/ Ceefax (68%), and national radio and tv (both 61%). The least used services were club radio (7%), the official website (at a relatively low 14% given the hopes invested in them), premium rate phone lines (18%, but, interestingly, thought by clubs to be one of the most popular services with fans) and information brochures (provided by 54% of clubs but used by just 4% of fans). Major differences existed between the services used by fans of lower division clubs and the top clubs. These were particularly pronounced in the use of the national and local news media. For example, national newspapers were cited by 94% of Premier League and Division 1 club fans, but by only 40% of lower division fans; national tv by 83% and 20% of premier and lower division fans respectively. Lower division fans used local papers more frequently (100% as against 61% of fans of top clubs), as well as local tv, Teletext/Ceefax and unofficial websites.

The average fan spends £1 to £5 per week on football information services; and 46% stated they were willing to pay for new or improved services. Teletext was rated by fans as the service giving best value for money (71% of fans thought so); match programmes were thought to be the most accurate (75%), followed by local papers and national radio (57%); local radio was thought the most current (96%). Least value for money were thought to be the premium rate phone lines (36%); somewhat disturbingly the least accurate were thought to be national newspapers (25%); and the least current were club magazines (28%).

Overall fans were fairly satisfied with the job their clubs are doing in providing information (*Table 8*). Half the fans thought their club was doing a good or excellent job. However over two-thirds of fans felt that they had little opportunity for feedback or a say in their club.

### Conclusions

There is more football information available than ever before, through a wider range of information channels. The demand for club in-

*Table 8. Satisfaction with clubs' information services*[1]

| Rating | Total (28) | Prem/Div 1 (18) | Div 2/3 (10) |
|---|---|---|---|
| Excellent | 7 | 11 | 0 |
| Good | 43 | 45 | 40 |
| Average | 43 | 45 | 40 |
| Poor | 4 | 0 | 10 |
| Very poor | 4 | 0 | 10 |

1. Expressed as a percentage.

formation is still rising, but that for general football information appears to be levelling out. In fact, 16% of supporters and many of the experts interviewed felt there is too much general football information. Major differences exist between the quantity and range of information and information services provided by and about the top clubs and those in the lower divisions. Demand for information at the lower level is not being met adequately and probably economics has something to do with it.

The growth in demand for football information and information services reflects the rise in the profile of the game in England since the late 1980s. Football is perceived to have overcome many of the image problems of the past. The game has made a conscious effort to appeal to a wider audience, and is no longer seen as the reserve of the working class male. Football clubs have attracted new supporters by improving facilities, comfort and safety at stadia, by introducing family enclosures, closed circuit tv to deter violence, and by mounting campaigns against swearing, racism and sexism inside grounds. It has become much more respectable to be a football supporter; in large part this has been due to the success of Nick Hornby's semi-autobiographical novel *Fever Pitch*. The introduction of live television coverage has also helped increase awareness of football, without the feared decrease in attendances at games. Football supporters have themselves helped to restore a positive image through concerted efforts to democratise the game and counter media negativity. In particular the growth of fanzines and independent supporters' associations has had a positive influence in raising the profile, popularity and respectability of English football.

Football is now a market with information as a product, supporters as consumers, and clubs as distinctive brands. There has been a movement from clubs relying financially on spectators paying at grounds, to new commercial areas of merchandise sales, sponsorship, corporate hosts and television rights. Football fans constitute a loyal breed of information consumer with a unique sense of brand loyalty. Football clubs are belatedly recognising that once a new supporter is made they will be a captive client for life, and clubs are marketing their brand on a national and, increasingly, global scale. Information services and information technology can help distribute the product to ever wider markets.

Football information services offer an opportunity for clubs to sustain their relationship with supporters, regardless of geographical dispersion, and can generate revenue directly as well as indirectly through merchandise sales. Global information services such as websites for English football clubs show heavy use from overseas fans. In turn, information and communication distribution channels see football as ideal content to sell their services and to provide the necessary audiences for advertisers. Football, therefore, has become a major item of content to push new and traditional information distribution channels to a global consumer market place. Investors and clubs are increasingly looking to exploit football information as valuable intellectual property and as a saleable commodity.

The survey shows that the main types of information offered by clubs to their supporters closely matches those sought by fans, namely facts and 'news' based information. This type of information is dominated by results, fixtures and match reports. This type of information is increasingly sought through continuously updated services such as Teletext, 24 hour sports radio and premium rate phone lines. As a result traditional media, especially newspapers, have been forced to produce more 'value added' types of information such as background features.

The influence of fanzines and democratic supporters' movements has given fans the opportunity to voice their feelings in football information services in a way that was previously lacking. However, there is evidence to suggest that football information is subject to increasing commercial pressures. Many fans believe that official information sources manipulate football information in their own interests, and both official and unofficial services deal in misinformation and ill-informed speculation. The result has been that while there is more information available, much is considered of poor quality and untrustworthy by football fans.

Four factors are going to prove critical to the future development of English football information services: ownership and control of information; the impact of new technology; the merging of the information and entertainment industries; and the widening information gap between Premier League clubs and the rest. Football clubs and authorities are beginning to assert their claim to ownership of every aspect of football which is under their control. As a result, football games, players and information are all seen as both product and intellectual property. The development of new mass communications technology will boost the range and total quantity of channels by which this content is delivered to football supporters. The key short term technologies will be digital television, the Internet, and Internet delivery via set top boxes for televisions. More football and football information may be good news for fans of the top clubs best placed to exploit these changes, but the gap is widening between those clubs and the rest outside the Premier League (perhaps even those outside the top ten or so of the Premier League).

At present the provision of football information is expanding to meet an existing and growing demand. There are, however, indications that this demand will not grow indefinitely before the football bubble bursts. Although developing quality and accessible services can generate income and help maintain competitiveness for clubs throughout the English leagues, as well as informing and entertaining fans, there is a danger that the market will become saturated and current levels of demand will not be maintained unless football supporters’ needs are adequately reflected in the provision of football information.

The exploitation of information as a potential source of revenue for clubs is still in its infancy, but information services have proliferated as clubs seek new ways to attract supporters and communicate with fans. Information itself is increasingly seen as a saleable commodity rather than as a public service to fans. Clubs now ‘sell’ information through magazines, information brochures, websites, phone lines, radio stations, museums, videos, and will soon do so through subscriptions to their own television channels[29]. They are also more adept at controlling and manipulating the information that leaks out through other channels.

The clubs and football authorities have been slow to react to previous possibilities in information and communication services. They resisted live TV coverage of football matches, yet many clubs are now reaping huge financial rewards. They thought that nobody would use the premium rate phone lines. That too has turned out to be a nice money earner. Having called it wrong so often they will surely not fail to recognise the enormous potential of the Internet, especially in regard to marketing themselves to distant supporters at home and abroad. Information management and football holding hands, who would have thought it five years ago, but that is the reality we face.

## References

1. Bradford, N. and Reeve, W. *Net profits*. (London: Fletcher Research, 1997.)
2. Wylie, M. Personal interview, 16th July 1997.
3. When Saturday Comes. *http://www.wsc.co.uk/wsc/*, 1997.
4. Shaw, P. *Whose game is it anyway?* (London: Argus, 1989.)
5. Robinson, J. *The best of football fanzines*. (London: Fanzines Publishing Co., 1989.)
6. Haynes, R. *The Football Imagination*. (Aldershot: Arena, 1995.)
7. Turner, R. *In your blood: Football culture in the late 1980s and early 1990s*. (London: Working Press, 1990.)
8. Sir Norman Chester Centre for Football Research ‘Fanzine survey findings.’ In *City Gent*. Bradford City FC Fanzine, Issue 25, September 1989.

9. Barrett, E. 'The rise and rise of the new fan.' *The Sunday Times*, 23rd March 1997, p.1.

10. Bishop, S. and Veljanovski, C. *Economics of league football.* (London: Case Associates, 1997.)

11. Varley, N. 'Mellor targets Sky squeeze on soccer fans.' *The Guardian*, 16th August 1997, p.3.

12. Moray Araujo, M. 'FourFourTwo'. *TV Times*. June 1996, pp.113-119.

13. Garrett, A. 'Premiership squares up for the big match over broadcasting bonanza.' *The Observer*, Business section, 27th July 1997, p.9.

14. Smith, G. Personal interview, 23rd September 1997.

15. Kelly, S. *Back page football: A century of newspaper coverage.* (London: Queen Anne Press, 1988.)

16. Walvin, J. *The people's game.* (London: Allen Lane, 1975.)

17. Hornby, N. *Fever pitch.* (London: Victor Gollancz, 1992.)

18. White, J. Personal interview, 6th October 1997.

19. Holding, I. Personal interview, 14th July 1997.

20. Fynn, A. and Guest, L. *Out of time: why football isn't working.* (London: Simon & Schuster, 1994.)

21. Barber, G. *Fan power and democracy.* Factsheet 7, (Sir Norman Chester Centre for Football Research, 1995.) *http://www.le.ac.uk/snccfr/research/fsheets/fofs7.html* (25/6/97).

22. Ryan, M. 'Internet thugs 'planning riot''. *The Daily Mirror*, 9th April 1997, p.33.

23. O'Reilly, D. 'Security fears spread as thugs turn high tech.' *The European*, 3rd April 1997, p.15.

24. Rose, P. Personal interview, 10th July 1997.

25. Herd, M. 'Bates's blueprint for the millennium.' *Evening Standard*, 24th July 1997, p.71.

26. Smith, G. 'Chelsea tune in with help from Sir Bob.' *The Daily Telegraph*, 11th February 1997, p.26.

27. Pennington, D. 'Leeds media plans.' *The Times*, 17th April 1997, p.31.

28. Buckingham, L. 'Digital tv set-top box row settled.' *The Guardian*, Finance and Economics pages, 27th September 1997, p.24.

29. Mulholland, J. 'United joins BSkyB in digital deal.' *The Guardian*, 25th September 1997, p.9.

# The News International Newspapers' intranet

*Gertrud Erbach and Lynda Iley*

*Reference Library, News International Newspapers Limited, 1 Virginia Street, London E1 9BD*
*Email: gertrud.erbach@newsint.co.uk*

**Abstract** *Since 1985, News International has maintained an in-house database which provides access to several daily and Sunday newspaper titles and has developed into a powerful research tool for librarians and journalists. A plan to provide a web-based front end to the Editorial Database led to the development of an Editorial Services Intranet which also makes available content produced by Reference Library staff. The material ranges from the databases such as the library's book catalogue and a record of all aviation disasters to lists of all IRA attacks in mainland Britain and ministerial resignations.*

## How it all started

Since 1985, News International has maintained an in-house database providing access to the text of *The Times* and *The Sunday Times*. Full text of Times Supplements was added in October 1994. From 1996, there has also been access to the text of *The Sun* and *News of the World*. In addition, News International exchanges text with several other newspaper proprietors. With the addition of these other daily and Sunday nationals, the Editorial Services Database has developed into a powerful research tool for librarians and journalists at News International.

The Editorial Services Database originally ran on Information Dimension's BASISK, but was migrated to BASISPlus. Migration was completed by the end of 1996. The front end for the new version was a result of cooperation between an experienced online searcher from the Reference Library and the Editorial Services Database Administrator. Their combined efforts resulted in a user friendly and flexible front end and led to increased use of the database. As part of the migration, it was planned that the new version should have a Windows and a character cell front end with different levels of use for different types of users. The Windows front end would have used BASIS Desktop software while the character cell front end was running using FQM procs.

Desktop was never introduced for this purpose as it did not provide all of the functionality requested. It had some compatibility problems with the network software being used and was difficult to customise. Many changes required reconfiguration of each user's PC, and licenses were relatively expensive. An alternative windows-type front end was still required.

In 1996, Information Dimensions released, and strongly promoted, their BASIS intranet Solutions WEBserver – enhanced versions of the BASIS products designed to support the use of WWW technologies in intranet applications (*www.idi.ocld.org; www.id-london.co.uk*). BASIS WEBserver was shipped in combination with the Netscape Commerce Server. BASIS Release 8 fully incorporated BASIS web products and included BASIS Document Manager; BASIS WEBserver Gateway and Net-scape Enterprise Server supporting document delivery across the web and CIC (Corporate Information Centre), a rapid prototyping framework for intranet development. This software seemed the obvious choice as it was developed by the suppliers of the database software.

## Defining the project

By the end of 1996, the largest part of the News International migration project was complete and the path was clear to implement the web-based front end. A project was set up with the primary purpose of providing intranet access to the Editorial Database. Netscape browser software was used initially but has now been replaced by Microsoft Explorer.

The project aim was to create an 'Editorial Services intranet' to provide:

- access to the Editorial Database through a user-friendly GUI front end;
- access to other relevant sources of infor-

mation – this would include material from the Reference Library as a minimum; information from other departments would be desirable;

- access for all PC users;
- compatibility with the new News International production system, Hermes and its DocCenter archival function.

Some expected benefits were:

- the software was relatively inexpensive and News International already had a site licence for browser software;
- ease of administration as configuration of each user's PC would not be necessary;
- intranet access would provide the user-friendly windows-type interface most people are becoming used to and little user training would be required;
- this interface could provide access not just to the database, but other useful information such as Reference Library materials;
- the information would be constantly updated;
- potential cost savings due to the reduced need for printed information;
- the potential for improving lines of communication within the company;
- potential for integration with the Hermes archival function.

## Establishing the project team

The project was driven by staff from two departments within Editorial Services, the Reference Library and the Information & Database Services Department. Additional help was provided by one member of the IT department. The team which came together to develop the intranet brought different skills to the project. The Database Administrator, the project leader, had BASIS and Unix skills; the Electronic Information Manager managed the content and the Internet Librarian had some design and HTML skills.

However, the intranet was a huge learning experience for everyone involved. Each member attended the Informed Business Services course 'Creating Web Pages'. The Database Administrator taught the Electronic Information Manager the Unix vi editor and directory structure on the server. This allowed the librarians freedom to create pages, add HTML and edit files. An HTML editor could have been used but this was decided against as editors tend to generate browser specific HTML and also can produce 'non-human readable' HTML; i.e. it can be very cluttered. A simple wordprocessing package could also have been used but vi proved quick and efficient.

The Electronic Information Manager and Internet Librarian attended the Information Dimensions course 'BASIS intranet DBA' to gain an insight into database administration and managing databases for the intranet.

## Developing and updating content for the Editorial Research intranet

Initially, content for the intranet was planned with editorial users in mind in line with company policy to increase direct access to information for journalists on their desktops. Direct access to the Editorial Database via a simple, intuitive front end, allowing journalists to search the Editorial Database without much need for training, was the first priority.

The Reference Library at News International is a very busy department. Operating 19 hours a day, seven days a week the department provides an information and research service for journalists from News International's four titles and corporate staff. Between 600 and 700 enquiries are handled on an average weekday. Much of the library's work is reactive – finding background material to breaking news stories for reporters or checking facts and figures for sub-editors. Certain types of enquiries are, however, recurring such as requests for the spelling of names, basic information on a country, the chronology of events of an ongoing story such as the Northern Ireland Peace Process or economic indicators. In anticipation of these recurring enquiries library staff over the years have compiled chronologies, lists and indices to make some of the information available quickly to allow for a more efficient service. Some information was available on Cardbox databases.

The intranet seemed the ideal medium to make these information resources directly available to journalists' desktops giving them speedier access to certain types of information and reducing pressure on the library staff.

### Editorial Database

The priority for the intranet was a search screen for the Editorial Database. The database has been split into two due to its sheer size. The first database stores data from 1985-1994 and the second data from 1995 to date.

Librarians access the database via a telnet session using the BASIS command language. It is very powerful but users need a high level of training. While the command language does not pose a problem for experienced researchers it can be difficult for end users. The web front end was designed so that anyone could use the database. Context sensitive help is provided on each screen in case there are problems.

The web front end uses a Perl CGI script which sits in front of the standard BASIS Webserver CGI binary. It uses cookies to keep the search terms available to the server when each request is performed and when refining searches. The search screens are generated dynamically by the Perl script depending on which database is being searched. Users on the intranet can choose between the two databases. It also performs manipulation of the search terms (dates and publications) so that they are understood by the standard BASIS Webserver gateway.

The database is made up of approximately 35 fields, some of which are used for searching. The web front end initially used only four of these fields, date, headline, byline and story. The default was to search all available titles but the user also had the choice of restricting a search to tabloids, broadsheets or supplements.

In January 1998 an advanced web front end was introduced which allowed users to choose individual titles and search using the headline and first two paragraphs of text only. The dates on the search screen appear as drop down menus. Headline, byline and story are text fields and users only need to enter words or phrases then click on search and they retrieve a list of headlines. The headline is then used as a link to the full text of the article. One advantage for the journalist is that they can take extracts from retrieved articles, edit or rewrite these extracts and paste them into their current stories rather than having to print or download into a file first.

Journalists prefer searching the Editorial Database via the web front end despite the fact that speed of retrieval is rather slower than through a telnet session.

### Reference Library databases

The next project phase aimed to make some of the Reference Library materials more widely available.

The Reference Library had used Cardbox Plus software to compile five small databases. These included a record of all aviation disasters and a record of murders of police officers in Britain. The team decided to combine these databases into one large BASIS database. The first step was to create a database definition list (ddl) or a set of fields which made up the database. There were a number of common fields for each of the small databases, e.g. name. A number of fields overlapped and some could be combined. There were also a number of unique fields for each database. One of these unique fields was used from each of the Cardbox databases to establish the view. For example, police force only appeared in the police murders database and this field identified records which should appear in the Police view of the database. It was decided that there should be one 'reflib' ddl and a number of different views so the user would see five 'different' databases even though there was only one underlying database. The five views were then made available through BASIS Desktop for staff to enter new records and make any amendments. Help sheets were written to ensure consistency of data input.

A new format was then designed on Cardbox Plus to enter the correct fields for the new BASIS Plus database and take into account the new and combined fields. The records were dumped into a text file out of Cardbox. This file was then transferred to the server using Onnet file transfer protocol (FTP). A copy of the file was made on the server and the database administrator split it using a script and ran a series of perl scripts to tidy the files, this ensured the data was displayed in the same format; e.g. on some of the databases block capitals had been used, a script replaced these with lower case except for names or first words of sentences.

The new files were then loaded using High Volume Upload (HVU) on BASIS. Records which failed validation, that is those which had incorrect date formats or had empty unique fields, were placed in separate files and passed to the Electronic Information Manager for correction. These were then copied to a retry directory and loaded.

Default BASIS Plus Webserver forms were generated using bwgenform – this provides default forms for searching and displaying using a browser. These were edited, for example the default search and help buttons were changed, the search forms appeared as a table, several superfluous buttons were removed, fields which were not used for searching were removed and the remaining fields were arranged in a logical order.

The same procedure was followed for the book catalogue.

New databases have been added over the past year including motor racing accidents and disasters. Work is ongoing on a country database with basic information on each country and a contacts database giving telephone numbers of organisations.

## Reference Library research material

Reference Library research projects ranged from the life and times of various celebrities to a list of all IRA attacks on mainland Britain, a history of the Peace Process in Northern Ireland, a complete list of all ministerial resignations and sackings since 1900 and various histories of the Royal Family. There were 62 of these, all available as Word documents. The files were saved as text files and then moved onto the server using FTP. HTML was added including appropriate links, for example a link from IRA Attacks to the Peace Process and vice versa.

In the past year around 15 new projects have been added including a history of the FA Cup and a complete guide to the World Cup. The World Cup pages were updated daily to give match results, scorers and attendances.

## Miscellaneous content

### *Conversion tables*

Metric/Imperial conversion tables are provided. There is some Java script which enables users to calculate distances (e.g. what is 565 kilometres in miles and vice versa), weights and measures.

### *Public libraries*

The Reference Library has a record of major public reference libraries holding copies of News International titles on microfilm or in hard copy, allowing the Library to direct callers enquiring about availability of newspaper material to their nearest public library. This was made available on the intranet for all departments who receive readers' requests.

### *Associations*

Journalists often need to ring trade associations to gain facts or soundbites for stories. Rather than ringing the library to gain a name and telephone number, a list of associations arranged by subject was made available. Therefore, if a story is being written on asthma and the journalist needs accurate numbers, they can look up the National Asthma Society without ringing the library and holding on in the queuing system.

### *Internet links*

Searching the Internet for accurate, useful information is still a time consuming task. The Internet Librarian originally spent some time trawling the web and finding such sites. A list of links was then produced which was divided by subject, e.g. Government, Entertainment etc.

### *Style guides*

Newspapers have their own in-house style guides which aim to provide writers and subeditors with a quick reference to points of grammar and spelling and to guide them through protocols in specialised areas where confusion may arise. *The Times Style Guide* was added to the intranet site and has since been updated, ensuring that the electronic Guide is always more current than is possible with the hard copy version.

The style guide is complemented by a list of frequently asked spellings not covered in the guide and generated by library staff.

## Maintaining and updating the content

Keeping content up to date is one of the most common problems identified for intranets[1].

There were two aspects of updating to consider. The first was that much of the initial content on the intranet would become unreliable without regular updates and would lose credibility with journalists.

It was unrealistic to expect one person to keep up with updating 70 research projects and 5 dynamic databases. Therefore it was decided to give an introduction to BASIS Desktop to each of the staff who had compiled the databases initially. They were then given the responsibility for keeping the databases up to date. The research projects are updated by one member of staff who has been allocated three hours each day to scan newspapers and update where relevant. Updating is fairly straightforward. The information is entered and automatically updated on the server.

Secondly, it quickly became apparent that without the introduction of new content the service would stagnate. Reference Library staff have continued to develop research topics and databases and these have been added. A 'What's New' section has been established on the Home Page to attract the users' attention.

## Promoting the intranet

Once the initial content was complete, the intranet was named Edse-Web (Editorial Services Web) and was successfully promoted within the company as a valuable research tool for journalists.

## Training

Edse-Web has been available since July 1997. At that time journalists at News International were still all using dumb terminals which did not support Internet/intranet access. In September 1997 *The Sunday Times* became the first title to move to the new NT environment; the other three News International titles (*Sun, Times, News of the World*) have since followed.

Journalists are busy people and are often reluctant to make time for training. The Reference Library's online training strategy acknowledges this and training is largely undertaken on an ad hoc basis. With this in mind, the intranet was specifically designed to be as simple as possible. There are help screens where appropriate. There are easy menus. All pages are displayed within a frame with a side menu to ensure quick, simple navigation. However, it is still useful to sit with users for 15 minutes and take them through what is available. Help sheets and leaflets detailing what is available via the intranet have also been distributed.

However, a more structured approach has been possible on *The Times*. The paper recently moved to new offices and on to an NT platform. As part of a compulsory programme of NT training, journalists were given an introduction to the intranet which has been followed up with further training by librarians. Although the training was minimal the benefits can be seen in increased and more effective use of the Editorial Database.

Library staff have also made training visits to the company's bureaux in Glasgow and Manchester, offering a flexible training timetable. Two trainers offered short training slots throughout the day. This allowed journalists the flexibility to be trained when they could find the time. A broad overview of the intranet can be provided in 15 minutes but once the journalists sit down and realise the benefits, they tend to stay for longer (sometimes an hour) asking questions and offering suggestions.

## Future developments

Content for Edse-Web will continue to be developed and managed by the Editorial Services Department. New databases are already under construction and Reference Library staff are working on information packages relating to forthcoming events. It is hoped that, as access to the intranet becomes more widely available, users will recognise its value to the organisation and submit their own contributions for inclusion. An email link invites users to send comments and suggestions.

## References

1. Morton, Ciaran. 'INTRANETS: Some Problems and Solutions.' *Managing Information*, 5 (4), May 1998, pp. 26-27.

**Subscriptions**

*Aslib Proceedings* is published ten times per annum and supplied to corporate members of Aslib. For 1999 the annual subscription rates are £170/$280 (£136/$224 to members of Aslib). For subscriptions and back issues please contact: Portland Press Ltd, Commerce Way, Whitehall Industrial Estate, Colchester CO2 8HP, United Kingdom. Tel: +44 (0) 1206 796 351. Fax: +44 (0) 1206 799 331.

---

Aslib, The Association for Information Management, is a world class corporate membership organisation with over 2000 members in some 70 countries. Aslib actively promotes best practice in the management of information resources. It lobbies on all aspects of the management of, and legislation concerning, information at local, national and international levels. Aslib provides consultancy and information services, professional development training, conferences, specialist recruitment, Internet products, and publishes primary and secondary journals, conference proceedings, directories and monographs.

Further information can be obtained from: Aslib, The Association for Information Management, Staple Hall, Stone House Court, London EC3A 7PB, UK. Tel: +44 (0) 171 903 0000. Fax: +44 (0) 171 903 0011. Email: aslib@aslib.co.uk *http://www.aslib.co.uk/*

---

Periodicals Postage Paid at Rahway NJ and at additional mailing offices.
**POSTMASTER:** Send address changes to PROCEEDINGS, c/o Mercury International, 365 Blair Road, Avenel NJ 07001

---

©1999 Aslib and contributors ISSN 0001 253X

ISSN 0001 253X
Volume 51
Number 3
March 1999

27|8|99
wcl

# Aslib Proceedings

## New information perspectives

### Deconstructing 'knowledge management'
David Streatfield and Tom Wilson

### Dumbing down: bottoming out?
Peter Cole

### United Kingdom regional newspapers on the World Wide Web
Kara Neilson and Peter Willett

### The changing information environment in the media: case study *The Guardian/Observer*
Helen Martin

Aslib THE ASSOCIATION FOR INFORMATION MANAGEMENT

*Aslib Proceedings* is a long-standing professional journal of international repute, which has a large and wide subscriber base. The journal brings currency, authority and accessibility to the reporting of current research, issues and debates in the broad area of information work. Above all the journal wishes to provide research and comment in a form that is easily and quickly understood: a fresh, rigorous, but unfussy, writing style is what is aimed for. The needs of busy practitioners (and academics) is very much in mind. Articles are now refereed to meet the growing need for accredited and authoritative information. To ensure that this does not result in an unacceptable loss of currency, articles are speedily refereed by the journal's large, eminent and multi-disciplinary Editorial Board (assembled for just such a purpose). In exceptional cases they may be referred to an external panel of consultants. The journal's monthly publishing pattern helps in maintaining the currency of reporting. In consequence the submission of research in progress articles is encouraged.

The boundaries of information work are rapidly changing and widening. *Aslib Proceedings* has a role to play in the repositioning of the information profession by encouraging submissions from information managers, providers, packagers and consumers of all kinds – and especially those from the related information fields of journalism, electronic publishing, communication and Internet studies. An interdisciplinary Editorial Board has been assembled to assist in the process.

Prospective authors should submit articles of 2,000-6,000 words in length. Articles should be accompanied by a 100-150 word informative abstract and the numerical form of referencing should be adopted. In style and tone, submissions should fall between that of the journal's sister publications: *Managing Information* and *Journal of Documentation*. Articles should be furnished by email or on disc to David Nicholas, Editor, Aslib Proceedings, Department of Information Science, City University, Northampton Square, London EC1V 0HB. Email: nicky@is.city.ac.uk. The Editor is happy to provide advice on an informal basis to potential authors.

**Editorial Board**

Dr David Nicholas, *Head of Department, Department of Information Science, City University;* John Akeroyd, *Director, Learning Resources, South Bank University;* Sarah Blair, *Head of Publications, Aslib;* Peter Chapman, *Information Manager, NewQuest North East;* Professor Peter Cole, *Head of Department, School of Journalism, University of Central Lancashire;* Dr Tom Dobrowolski, *Senior Lecturer, Institute of Information Science and the Book, University of Warsaw;* Harry East, *Senior Research Fellow, Department of Information Science, City University;* Mike Hyman, *Managing Director, Head Software International;* Robin Hunt, *Creative Director, AreHause;* Nat Lievesley, *Network Manager, Centre for Policy on Ageing;* Helen Martin, *Information Manager, Research and Information Unit, The Guardian*; Dr Ian Rowlands, *Lecturer, Department of Information Science, City University;* Richard Withey, *Director, New Media, News International.*

**Editorial Team**

*Editor:* David Nicholas
*Deputy Editor:* Ian Rowlands
*Editorial Manager:* Kay Nicholas
*Editorial Assistants:* Peter Williams (City University), Colette Coles (City University)

# CONTENTS

67-71 Deconstructing 'knowledge management'
David Streatfield and Tom Wilson

72-77 Dumbing down: bottoming out?
Peter Cole

78-90 United Kingdom regional newspapers on the World Wide Web
Kara Neilson and Peter Willett

91-96 The changing information environment in the media: case study *The Guardian/Observer*
Helen Martin

# Deconstructing 'knowledge management'

*David Streatfield*
*Principal, Information Management Associates, 28 Albion Road, Twickenham TW2 6QZ*
*Email: Streatfield@compuserve.com*

*Tom Wilson*
*Professor, Department of Information Studies, University of Sheffield, Western Bank, Sheffield S10 2TN. Email: t.d.wilson@sheffield.ac.uk*

## Prologue: a fairy tale

*Once upon a time there were four Goblins who were, of course, chartered accountants. They became tired of counting hoards of gold for the Gnomes so they sought help from the Wizard of Ecognomeics. 'We want our own treasure chests,' they pleaded. 'That's easy,' said the Wizard 'Human beings have the idea that information technology will solve all their work problems. I'll turn you into Management Consultancy Goblins. Then you can each take one idea from the Academic Ivory Tower and that should make you piles of gold.'*

*After 5,000 days the four met up at the International Goblin Warming Conference and compared experiences. 'I chose Corporate Management Information Systems,' said Goblin 1, 'and I was selling loads of mainframes. Then the Gnome Computers Wizard invented the personal computer and everyone started networking instead.' Goblin 2 had taken the software route. 'My Computer-Assisted Management Systems were selling like hot cakes until somebody noticed that they were only doing administration. Then it all went pear-shaped.' 'I chose fear,' said Goblin 3, 'All the companies were latching on to my 'Information for Competitive Advantage' until someone asked 'What happens when everyone is doing it?'*

*The fourth Goblin adjusted position on his platinum toadstool, glanced at his Cartier watch and asked, 'Have you sad Goblins ever heard of my Knowledge Management?'*

## Introduction – the rise and rise of knowledge management

Is knowledge management just another fad? No, in the sense that people are trying to make sense of the idea of managing information as a strategic resource or asset on a par with managing people, material resources or money. Yes, in that over-simplistic use of the concept of knowledge has led to overblown claims by IT advocates and management consultants that all knowledge lies within the domain of the latest (or next) generation of organisational computers.

This oversimplification is common in the UK library and information press and has resulted in attempts to make a too-easy synthesis of the roles of information manager and knowledge manager[1]. Fortunately, there is more clarity in the body of recent US work which has emerged in response to the 'pre-packaged and taken-for-granted' interpretation of knowledge and its management by IT systems management consultants when advancing the concept.

The idea that knowledge management is just the latest management consultancy fashion statement, following hard on the heels of Total Quality Management and Business Process Re-engineering is quite widely perceived – even amongst some management consultants. Speaking at the 1997 *International On-Line Meeting*, Clive Holtham of City University Business School demonstrated that management consultancy fads have a duration of about three years and forecast that knowledge management as *the* management elixir was likely to follow this pattern. At the same event, David Best (then a partner in Deloitte & Touche Ross Consulting Group), noted that a great deal of the discussion on knowledge management regressed into talk about traditional library and information management activities.

## The knowledge management agenda

In presenting an overview of 'Knowledge management in enquiring organisations', based on a substantial literature review, Malhotra[2] characterised 'most extant formulations of IT enabled knowledge management' as ignoring humans in the process of knowledge creation. Similarly, Davenport[3] complained that the IT proponents were advancing a static syntactic notion of knowledge and disregarding how people in organisations actually go about acquiring, sharing and creating knowledge. Unfortunately this trend has continued[4] but the tide appears to be turning in favour of a more holistic organisational view. This shift in view has come about largely because IT investment as a path to organisational success has been exposed as a chimera (notably by Paul Strassman, whose industry-wide analyses of IT investment concluded that there is no relationship *whatsoever* between computer expenditure and company performance)[5].

The idea of information as a resource to be managed strategically, alongside the management of people and material resources, is not new. Management writers from Mintzberg[6] onwards have been urging Chief Executives and other senior managers to consider organisational information more holistically. This led to exploration of a series of useful and applicable concepts such as the 'learning organisation'[7] and the 'know-how company'[8].

Despite these arguments, the knowledge management metaphor remains an enticing one, whether the intention is to promote the sale of IT systems, advance the position of information workers to the status of knowledge managers[9], create a new market arena for specialist business consultants by coupling the concept with the notion of competitive advantage[10], or from a very different perspective, to provide more effective means of 'managing intellectual capital' and creating knowledge through social interaction[11,12,13] because 'only human beings can take the central role in knowledge creation'[14].

## The knowledge awareness perspective in managing people

Since we are evidently unsympathetic to misusing the terms knowledge and management in combination, what do we propose instead? In his seminal article 'Deciphering the knowledge management hype', Malhotra[15] offered his own definition:

> 'Knowledge Management caters to the critical issues of organisational adaption, survival and competition in face of increasingly discontinuous environmental change. Essentially, it embodies organisational processes that seek synergistic combination of data and information processing capacity of information technologies, and the creative and innovative capacity of human beings.'

This definition was in part advanced to challenge the inflated claims associated with new techniques 'such as datamining, intranets, videoconferencing and webcasting... as panacea for the business challenges of the knowledge era'. It will also help as a starting point for positioning the information professional within the knowledge management arena.

Under the general heading of knowledge management, it may be useful to distinguish between:

- the concept of knowledge as an individual or social construct (which becomes heavily problematic if the organisation seriously sets out to 'manage' what is in people's heads!);
- the adaptable organisational environment in which people are encouraged to learn and to share their knowledge; and
- those aspects of organisational 'knowledge' or 'intelligence' (in the police or military sense) that are amenable to organisation, re-engineering, planning and development.

From an organisational perspective, there is much merit in encouraging an adaptable learning environment and in focussing on identifying and gathering organisational intelligence in a strategic manner. There is also value to the organisation in encouraging the effective transfer of expertise, know-how and knowledge. These ideas might be summarised as a 'knowledge awareness strategy'.

The essence of knowledge awareness management is that *all* key staff should be treated as information managers, even if one of them assumes overall responsibility for the strategy. This does not disqualify the information professional from taking a key role in the organisational attempt at developing knowledge awareness strategies; nor does it assume any pre-defined role as proper for the information professional.

## The role of the information professional

What, then, is the role of the information professional in an organisation trying to adopt a knowledge awareness perspective? The basic role itself is unchanged: the organisation will still expect the information professional to acquire and organise information, make it available for use, disseminate it and, ultimately, dispose of information resources in all media and formats. A knowledge awareness perspective in the organisation calls for the information professional to be more systematic in tracking existing organisational information strengths and this will probably lead to developing new information resources relating to the available in-house expertise.

There is, of course, nothing remotely new about tracking the expertise of staff. Organisational skills inventories or indexes of expertise have been around for at least 35 years[16]. They now appear to be making a come-back as the 'knowledge index', but, in form and intention, this is the same entity. Other well-known tools that have prospered at various times and then either simply entered the general toolkit or have disappeared include the diary of conferences and exhibitions attended by members of the organisation, with the information service taking responsibility for storing and disseminating the materials acquired during visits.

Senior managers who take a knowledge awareness perspective will also expect their information professionals to be proactive in gleaning organisational intelligence and to directly serve their needs. In a recent consultancy survey of further education college principals and their deputies[17] it was possible to identify a few colleges where the senior managers were really 'thinking strategic' about information. One characteristic of each of these colleges was that they wanted their librarian/learning resources manager to be a full participant in the senior management information exchange network that the interviews were piloting. None of the other colleges could see any relevance in involving the librarian, whose job was defined in terms of support for students and teaching staff. We have elsewhere noticed the tendency for organisational information services to concentrate their active services on headquarters-based middle managers and to offer their routine dissemination activities below that level – unless senior management insists on a more strategic focus. In the university world, for example, a study revealed that most university librarians in the sample did not believe that the library had a role to play in serving the *management* information needs of academic heads of departments[18].

How can information professionals adopt a more strategic organisational view of their work? Again we are not considering new concepts. One way to monitor what is really happening is through what Peters and Waterman labelled (belatedly) as 'management by walking about'. When one of the authors worked in industry he was advised by his (internationally renowned) Director to spend a third of his time walking around the organisation talking to people about what they were doing. Gathering information on what was going on in different departments (that frequently did not talk to one another) paid off: on one occasion, hundreds of thousands of pounds were saved when it was discovered that two different departments were working on the same problem in ignorance of each other. That was in 1960 – 39 years ago, when the only computer in the company occupied the entire top floor and when knowledge management had never been brooked!

However, fully to engage with the 'learning organisation' requires that the librarian or information manager finds a new locus within the organisation – ideally, if information professions are to play a full part in such organisations, they should be located in the Chief Executive's office. They should form part of a team devoted to acquiring external information of value to organisational strategy and establishing systems to ensure that the knowledge held in people's heads can be tapped when it is needed. Merely to leap into the 'knowledge management' balloon without being in a position fully to deliver will simply mean being first *out* of the balloon when ballast is jettisoned.

## Conclusion

> 'Knowledge is not a 'thing' that can be 'managed'. It is a capacity of people and communities, continuously generated and renewed in their conversation, to meet new challenges and opportunities. People responsible for knowledge value creation can be inspired and supported, but they cannot be 'managed' as people were managed in the industrial era, as mere extensions of the machinery.'
>
> George Por 'Why knowledge management is an oxymoron' Community Intelligence Labs. 1997.

George Por's statement cuts away the hype of the knowledge management movement and demonstrates that the idea enshrouds two stronger concepts: the management of information and the management of people. We cannot manage knowledge directly – we can only manage information about the knowledge possessed by people in organisations. Even then, the information to be managed is necessarily incomplete because the boundaries of personal knowledge are fuzzy and continually changing as individuals get out of touch or extend their knowledge.

Adopting a knowledge awareness strategy requires all managers to work with their staff to ensure that people can 'learn on the job' and that information about their expertise and know-how is shared. The essential role for the information professional within an organisation that aspires to enhance its knowledge awareness is – managing information. Senior management may ask the information professional to go further than previously in helping to encourage the effective transfer of expertise, know-how and knowledge within the organisation but these are not new roles for the proactive information professional.

It is 30 years since Peter Drucker[19] began to write about the challenge to organisations presented by 'the knowledge economy' and it may well be claimed that we have been slow to respond to this challenge. Enhancing the knowledge awareness perspective of an organisation demands a shift in culture in the direction of greater openness, with the attendant risks and costs. There is no 'quick fix' for the organisation: simply to proclaim the Quest for the Knowledge Management Grail (even if this includes recruiting a Chief Knowledge Officer) will make no difference in meeting Drucker's challenge.

## References

1. See, for example, the publicity issued by Aslib for Sylvia Webb's *Knowledge management: linchpin of change* published in Autumn 1998; or Stewart, C. 'Knowledge management: building the desktop environment.' *Library Association Record* 100 (4), 1998, p.192.

2. Malhotra, Y. 'Knowledge management in enquiring organisations.' in *Proceedings of 3rd Americas Conference on Information Systems* Indianapolis, IN, 15-17 August 1997; pp.293-295.

3. Davenport, T.H. and Prusak, L. *Working knowledge: how organisations manage what they know.* (Boston, MA: Harvard Business School Press,1997.)

4. See, for example, Liebowitz, J. and Wilcox, L.C. (editors) *Knowledge management and its integrative elements.* (Princetown: CRC, 1997.)

5. Strassman, P. *The squandered computer.* Information Economics Press 1997.

6. Mintzberg, H. 'Planning on the left side and managing on the right.' *Harvard Business Review*, July-August 1976, pp.49-58.

7. Pedler, M., Boydell T. and Burgoyne, J. *The learning company.* (Maidenhead: McGraw-Hill, 1991.)

8. Sveiby, K.E. and Lloyd, T. *Managing knowhow: add value... by valuing creativity.* (London: Bloomsbury, 1987.)

9. Fisher, A. 'Knowledge management: so what is the big fuss about?' *Library Association Record* 100 (4), April 1998, pp.190-1.

10. Such as Andersen's 'Knowledge Xchange, or CAP Gemini's 'Knowledge galaxy'.

11. Sveiby, K.E. *The new organisational wealth: managing and measuring knowledge-based assets.* (San Francisco, CA: Berrett-Koehler, 1997.)

12. Sanchez, R. and Heene, A. *Strategic learning and knowledge management.* (London: Wiley, 1997.)

13. Stewart, T.A. *Intellectual capital: the new wealth of organisations.* (NY: Doubleday, 1997.)

14. Nonaka, I. and Takeuchi, H. *The knowledge-creating company.* (Oxford: OUP, 1995.)

15. Malhotra, Y. ‘Deciphering the knowledge management hype.’ *Journal for Quality and Participation,* July/August 1998. [Special issue on Learning and Information Management]

16. See, for example, Maozell, R.E., *et al.* ‘An index of skills for research and development personnel.’ *Research Management* 6 (3), 1963, pp.209-210; Bronstein, R.J. ‘Setting up a skills inventory: how to do it on a shoestring.’ *Personnel,* March/April 1965, pp.66-67.

17. Streatfield, D.R. *Report to the Further Education Development Agency on an information exchange service for senior college managers.* (London: Information Management Associates, 1995.)

18. Greene, F., Loughridge, F.B. and Wilson, T.D. *The management information needs of academic heads of departments in universities.* (Sheffield: University of Sheffield, Department of Information Studies, 1996.) (BLR&D Report 6252)

19. Drucker, P. *The age of discontinuity: guidelines to our changing society.* (London: Heinemann, 1969.)

# Dumbing down: bottoming out?

*Peter Cole*
*Professor of Journalism, School of Journalism, University of Central Lancashire, Preston PR1 2HE. Email: p.g.cole@uclan.ac.uk*

**Abstract** *Weighs up whether a dumbing down has taken place in the British tabloid and broadsheet press. Material is presented from* The Sun, The Daily Telegraph *and* The Times, *and also from a number of television channels for comparison. Provides evidence which suggests that, while there has been changes in news reporting it is not strictly one way and down; indeed, there is some evidence to indicate an upmarketing amongst the tabloid press and television.*

## Introduction

Tuesday 11 August 1998. Front page of a major British newspaper. Two trails for the good things inside the newspaper, to tempt the casual reader at the newsstand. Full colour, four columns for each. Trail – journalists call them 'blurbs' – one: 'Anatomy of a power wedding.' Trail two: 'What Anthea did next.' The latter, on which I will concentrate, shows a horizontal photograph of one Anthea Turner, naked, accompanied by a python lying alongside her and shielding one bit of her person which would otherwise have been seen. The picture is completed by the symbolic presence, beside Anthea's elbow, of an apple. We are referred to 'News' (sic), page three. Just before we set off in that direction, let us note the words in large red type just above this provocative photograph. They read: 'Britain's biggest-selling quality daily.'

And so to page three, a significant page in any newspaper. Conventional wisdom has it that page one often takes care of itself, the news agenda dictating the prominence of most stories. Many readers turn quickly to page three, a page on which the editorial values of the newspaper are often laid bare, in this case literally. *The Sun*, which this paper is not, has put the phrase 'page three' into the language. Here we find the same picture that appeared on the front, this time larger, much larger. 14.5 cms by 35 cms, to be precise. Beneath a strapline – the paper puts them on each page to identify the subject area for that page – with the one word 'News'. And a headline reading 'Lovelorn Anthea grins and bares it.'.

Ms Turner is one of that new breed of media celebrities, most often found in front of a weather map or on breakfast television, who are famous for being not very famous. Her career progressed from Blue Peter to presenting the National Lottery show on Saturday nights. She achieved notoriety for leaving her husband and manager for a wealthy married businessman whose wife applied much guile and media help to getting him back. She succeeded. Ms Turner, distraught, turned to *Tatler* magazine and the aforementioned python, and together they made a cover story. This was duly recycled all over the newspapers. The version described above appeared in *The Daily Telegraph*, Britain's biggest-selling quality daily. Traditionally *The Daily Telegraph* is read by Conservative-voting middle England, by colonels and bank managers, and it has always numbered 'disgusted, Tunbridge Wells' among its readers. Whether he has heard of Anthea Turner is open to conjecture, as is whether he is prurient as well as disgusted. *The Daily Telegraph* has always used page three for the more salacious court cases, often giving more detail than more overtly salacious newspapers, but it is not often it resorts to such display as Anthea and the python.

This will have given further ammunition to those who argue that the British media is 'dumbing down'. A pause for definition. Dumbing down is used to mean making more populist, less cerebral. In terms of media, print or electronic, it means to include within an upmarket product (e.g. a broadsheet newspaper) content deemed (by the elite) inappropriate for an intellectual, social or cultural elite. Popular culture, mass media celebrities, soap opera, the sexual peccadilloes of the rich and famous, all are seen as representing dumbing down when included in upmarket media products. Those who use it always apologise for 'that horrible American expression'. It seems no more horrible than many clichés, and like many clichés it is a useful descriptor, imported or not. It is applied most often to the broadsheet, or 'quality', press (*The Times*, *The Daily Telegraph*, *The Guardian*, *The Independent*), BBC radio, and BBC and independent terrestrial television. The popular synonym is 'down-marketing'.

So if, as it did on Sunday 20 September 1998, Channel 4 gives an hour of its evening schedule to the attempts by a (genuine this time) page three girl, Melinda Messenger, to break into Hollywood, then that is dumbing down. In this case it started long ago in the quality broadsheets. In early 1997, when Melinda, whose ample attributes have some artificial content, first broke into tabloid toplessdom, the broadsheets were happy to report the fact, however bemused their readers.

> 'It is being hailed as one of the most meteoric rises in modelling history. A young woman from Swindon, who until this week appeared on bus-shelter poster ads for double glazing, is preparing today to fly to Los Angeles to be photographed alongside Hollywood's most famous swimsuit queen, Pamela Anderson.' That was page three – Home News (sic) – of *The Times*.
>
> 'The Sun is thought to be planning to anoint her as the new Pamela Anderson with a photo shoot on Santa Monica beach. Miss Messenger, who has 11 O-levels, carries a heavy burden – no, this is not a bosoms joke – she is to be the saviour of page 3.' *The Daily Telegraph*[1].

Dumbing down? There are other, less anatomical, examples. At random: when, during the last general election campaign, Tony Blair made 10 hand-written pledges to the nation, dashed off on his Islington patio, the quality press seemed to give more attention to analysing the handwriting than analysing the promises. *The Guardian* put the handwriting analysis on page one. Back to sexual matters. When Hugh Grant, British star of that most successful British film *Four Weddings and a Funeral*, was caught in an embarrassing situation by the Los Angeles police, the four British broadsheets gave a total of 1752 column inches to reporting and analysing this piece of news.

## Dumbing down – case proven?

It may seem from all this that the case that the British quality press has dumbed down is proven. Not necessarily. Certainly the examples given would not have appeared in the broadsheet press at one time, not so long ago. But that does not prove that overall, day by day, we have witnessed a downmarketing of the upmarket press. Certainly this is argued by some, and argued vigorously, but there are other factors at play, to which we shall come. Let us first consider the arguments that are made. Anthony Sampson, author of *Anatomy of Britain* and its various subsequent editions, and associated with the *The Observer* for many years, wrote in *British Journalism Review* in 1996 of the 'striking change in the character of the broadsheets or quality newspapers'. The frontier between 'quality' and 'popular' newspapers had virtually disappeared since the 1980s, he argued.

> 'It has led to less choice. In particular the high ground of serious reporting, investigating and foreign coverage has been vacated. Instead there has been an explosion of columns providing comment without facts, discussing friends, parties or other journalists. Newspapers are much less about history, more about conversation of the most basic kind: 'What did you do today?' Columnists tell us what happened to them on their way to Sainsbury's, what their children did at school, how they enjoyed their holidays. One wonders what there is left for these people to talk about at home.'

Sampson is in no doubt that these changes are real and for the worse, looking back to a time when 'quality' meant 'serious'.

> 'Does the triviality of the media really matter? Why shouldn't they entertain, if that is what people want? It does matter, above all because it affects the

> whole political process. A mature democracy depends on having an educated electorate, informed and connected through Parliament. But as the media have become more pervasive and more entertaining, Parliament itself is becoming marginalised in the national debate. Debates on television have increasingly taken over in public interest from Westminster. At best they can be much sharper than Parliament's, with fiercer questions, sometimes backed up by filmed interviews and evidence. They can sometimes detect groundswells of public opinion and revolt before they reach MPs' attention. Yet television debates provide no real answers, no solutions, no follow-up legislation. They are most exciting when they are most negative, like hecklers. And they increasingly eclipse the discussions in Parliament itself.'[2]

More recently the distinguished journalist Ian Jack has written on the subject. Jack was editor of the *Independent on Sunday*, and then editor of *Granta*. In a forward to the *Granta Book of Reportage*[3], Jack laments the passing of the tradition of straight reporting, famously espoused by the great American journalist Martha Gellhorn who died in 1998. Jack defines good reporting as describing a situation with honesty, exactness and clarity, delving into the questions who, what, when, why and how, without losing sight of the narrative. 'Newspapers and news organisations', writes Jack,

> 'are now part of the 'media industry' and a subdivision of showbusiness. The old distinctions between the serious and the frivolous – tabloid, middle market, broadsheet – have largely broken down. Meanwhile the craft of what the late Martha Gellhorn called 'serious, careful, honest journalism' has entered its own small crisis. Newspapers cutting their costs to meet reductions in cover price revenue could afford fewer reporters. Executives did sums.'

Jack says later in the forward:

> 'Reporting is in trouble more generally. Something in the climate mitigates against those Gellhorn words 'honest, careful, thoughtful'. Today a spectre haunts the editorial floor – the spectre of reader's boredom, the viewer's lassitude. If customers are to stay with the product, they need, or are thought to need, a diet of surprise, pace, cuts-to-the-chase, playfulness, provocation, drama, 'human interest'. ... Britain has developed a singular sort of media culture which places a high premium on excitement, controversy and sentimentality, in which information takes second place to the opinions it arouses.'

Nick Cohen, an *Observer* columnist writing in the *New Statesman* in May 1998, also attacks the column culture identified by both Sampson and Jack. In the 1950s and 1960s, he says, the craft of reporting commanded respect, and reporting was the best method of getting under the skin of a society, if you wished to be radical, or of finding something more interesting to write about than yourself and your friends, if you didn't. 'Today, if you want to get ahead, get a column, preferably with a blurred picture of your unprepossessing face at the top of it.'[4]

But let us take the debate to those who might reasonably be regarded as contributing to the dumbing down, editors of the quality broadsheets. At one of the prestigious Ditchley conferences two years ago, where the great and the good meet to talk off the record about matters of concern to them, and possibly a wider world, dumbing down came on to the agenda. Both Alan Rusbridger, editor of *The Guardian*, and Peter Stothard, editor of *The Times*, spoke on the subject. Both are relatively young editors; both have to deal day by day with practice not theory; they have to sell newspapers, produce newspapers people want to buy. They have to think, therefore, about content. Rusbridger started by thinking about young readers, always an editor's, and a circulation manager's, and an advertising director's obsession. Young readers, particularly if they have successful, high-earning careers ahead of them, are the ones all of these executives want. They consume more. They have longer to live. Rusbridger, tongue a little in cheek, described them thus:

> 'They don't read; they're not interested in politics or current affairs; they don't go to the opera or the theatre. They don't like heavyweight articles asking whither Nato. They're obsessed with American culture. Even the clever ones would rather read about The Simpsons and The Spice Girls than Schnittke or Stoppard. They're post-modern, post-serious, post-literate ... and post-broadsheet. The audience for serious journalism is, in other words, disappearing.'[5]

But there has been research carried out by the reputable MORI polling organisation showing that the situation is not as bad as Rusbridger described[6]. Two out of three of the British public regularly (defined as three or four times a week) read a national daily paper. The figure 20 years ago was 78%. Today 64% of young people – 15 to 25 year olds – regularly read a daily paper, just two per cent less than the population as a whole, and indeed a bit higher than the next two age groups, 25 to 34s and 35 to 44s. Of these young people, 15% read broadsheets. That upmarket figure has risen a little over the last 20 years. And what are these young people reading in their broadsheet papers? Mori asked questions about a range of subject areas. More young people read local and regional news than the population as a whole, and they were as interested in international news as they were in national news.

Rusbridger again:

> 'You can't put your head in the sand and just talk to graduates from the older universities. And you can't just shove in columns of text these days and expect people to read it. There are too many other media competing for people's attention. This is not C.P. Snow land where there's high culture and low culture. Most of our readers have the capacity to think more broadly than that. They want to know about the single currency and they would feel cheated if *The Guardian* didn't give that. But they also want to know about Liam Gallagher and Patsy Kensit. It's very dangerous if you get into the mindset that there are tabloid and broadsheet subjects. What there has to be is a unity of tone, and that's how you define broadsheet values.'[5]

Let us look at some data prepared for *The Times* by Glen Owen and reported by Brian MacArthur in his column in that paper[7]. The research looked at *The Times*, *The Daily Telegraph*, *The Daily Mail* and *The Sun* – we will take only the first two here – comparing content of issues from the first week of November in 1966, 1976, 1986 and 1996. First size. *The Daily Telegraph*, in those November weeks over the four decades, moved from 24 pages in 1966, through 32 in 1976, 36 in 1986 to 48 in 1996. Corresponding figures for *The Times*: 20, 24, 48, 48. On Saturdays the picture is much exaggerated. Back in 1966 the Saturday edition of the daily broadsheet was usually the worst seller of the week, and consequently a thinner product. Now Saturday papers are as bulky as Sundays, and sell more copies than any other day of the week. In 1996 *The Daily Telegraph* on Saturday contained 110 pages, plus a magazine, plus a television listings magazine, plus *Young Telegraph*. *The Times* ran to 116 pages, plus its own magazines/supplements.

But what about content? Design changes, and the modern broadsheet is a far more accessible product than its forbear of 30 years ago. Pictures are bigger, layout is brasher, packaging is given greater emphasis. So there are fewer words on each page. But the comparison was based on a true comparison of word count. Home news in 1966: *The Daily Telegraph* carried 5000 words of general reports and 3000 words of political reports. A total of 8000 words. *The Times* carried 5000 words of general reports and 7000 words of political. In all 12,000. In 1996 *The Daily Telegraph* carried about 15,000 words of home news with political news incorporated in the general run. *The Times* carried 13,000 words of home news, with the political content well down at under 3000 words. Foreign news: in 1966 *The Telegraph* was carrying nearly 5000 words, *The Times* nearly 8000. There was a big increase in both papers in the eighties, but by 1996 the figure was much closer to the 1966 level, 5000 words in *The*

*Telegraph*, 7000 in *The Times*. The space given to sport has grown massively in both papers. While it varies according to the day of the week, with special sections on Mondays and Saturdays, it is fair to compare a big sports day, Monday. In 1966 *The Daily Telegraph* ran 20,000 words of sport and *The Times* 16,000. In 1996 the figures were 40,000 for *The Telegraph* and 50,000 for *The Times*. They will be bigger still today.

We can see from that data that quantitatively the number of words given to those areas defined by Anthony Sampson as describing how upmarket a newspaper is have been maintained. With the exception of parliamentary coverage. Until not many years ago a distinctive feature of the broadsheet press was the space it gave to words spoken in the chamber of the House of Commons, the *Hansard* role. No spin, no interpretation, simply the words, and every name followed by brackets containing party and constituency. Why was it done? Even then it was accepted that few people read it. But the politicians 'demanded' it. Broadsheet papers saw themselves as 'papers of record'. At this time, and we are talking no more than 20 years ago, it was far more important who read *The Times* than how many read *The Times*, the 'top people's newspaper', as it thought of itself then. The beginnings of the non-deferential age, the decline in respect for politicians, the advent of Parliamentary broadcasting, all sounded the death knell for the reporting of debates. Non-establishment editors, themselves the product of the non-deferential sixties, recognised that their papers need not sell so few copies, could be allowed to be interesting, could broaden the scope of what they covered.

Simon Jenkins, a pretty establishment figure by modern standards, was editor of *The Times* when it dropped its parliamentary pages.

> 'I couldn't find anybody who read it, apart from MPs. We are not there to provide a public service for a particular profession, or for that matter for a particular legislative chamber. Newspapers are about providing people with news.'

## Upmarketing

Interestingly, just recently we have seen the first signs of upmarketing by the sector of the newspaper market few would have considered had any potential for dumbing down: the national tabloids, or 'red-tops' as they are often known – *The Sun* and *The Mirror*. This sector has seen the most significant fall in sales over the last year or so, and since these two newspapers sell more than six million copies a day, small percentage declines in sale mean large numbers of copies, with serious commercial implications for advertising as well as sales revenue. Both Rupert Murdoch's News International and David Montgomery's Mirror Group have addressed the problem by considering whether their newspapers were too downmarket. They have looked at the middle market and seen that *The Mail* continues to go from strength to strength, to add sales both during the week and on Sundays. Could it be that red-top readers are moving upmarket? Murdoch appointed a new editor at *The Sun*, David Yelland, and he is interpreting his brief as a nudge upmarket. The topless models are still there, but less prominent and displayed with less confidence. Issues are addressed. *The Mirror*, too locked into New Labour, is nevertheless moving away from its inevitably unsuccessful attempts to beat *The Sun* at its own game, and again is addressing issues of relevance to its readers instead of depending entirely on a diet of soap and celebrity. While a long way from *The Mirror* of the sixties, which took its 'educating and informing' role more seriously than is customary in the tabloid market today, it appears to be reversing the dumbing down tendency.

There are the first signs, too, of a similar change of direction in television news. Researchers Steven Barnett, of the University of Westminster, and Ivor Graber, of Goldsmiths College, this year examined one week of early evening news bulletins on BBC1, ITV, BBC Radio 4 and Sky News[8]. They were repeating an exercise carried out five years previously, and were therefore able to draw some comparisons. They categorised stories as 'serious' – politics, business, economics, foreign – or 'tabloid' – showbusiness, sport, crime and royal matters. In 1993 they found that ITV was dumbing down, and there was a clear divergence of news values between ITV and BBC1. ITV news content was 60% tabloid, BBC1's was 41%. In 1998 a different picture emerged, with significantly less airtime being devoted to trivia on both channels. In the same week five years later – and the researchers emphasise that it is only a week, and the 1998 week contained some big serious stories such as Hurricane Mitch, the Iraq crisis and the American midterm elections – ITV's 'tabloid' content was 35%, BBC1's 26%. Not only was the trivia content down for both, but the gap between the two channels had narrowed from 19% to nine per cent.

Just straws in the wind, perhaps, and much more detailed research will be required. But has dumbing down bottomed out?

## References

1. Boshoff, A. 'Lovelorn Anthea grins and bares it.' *The Daily Telegraph*, 11 August 1998, p.3.

2. Sampson, A. 'The crisis at the heart of our media.' *British Journalism Review* 7 (3), 1996, p.42.

3. Jack, I. Introduction to *Granta Book of Reportage*. (Granta, 1998.)

4. Cohen, N. 'The death of news.' *New Statesman*, 22 May 1998, p.18.

5. Rusbridger, A. *Ditchley Conference Papers*, February 1997(unpublished).

6. Mori. *Survey of newspaper readers.* (Mori, 1996.)

7. MacArthur, B. 'Content analysis carried out at *The Times* by Glen Owen.' *The Times*, February 1997.

8. Gaber, I. and Barnett, S. 'And now for the serious news.' *The Times,* 27 November 1998, p.45.

# United Kingdom regional newspapers on the World Wide Web

*Kara Neilson and Peter Willett*

*Department of Information Studies, University of Sheffield, Western Bank, Sheffield S10 2TN*
*Email: p.willett@sheffield.ac.uk*

**Abstract** *This paper discusses the ways in which regional newspapers in the UK have reacted to the challenges posed by the Internet. A detailed evaluation of the web sites for 25 regional newspapers demonstrated clear audience trends and a certain consensus of styles and features. The level of commercial activity in the form of advertising was found to be generally low, and the strategies to promote local business involvement varied. Nine of these 25 newspapers responded to a questionnaire that sought to gauge levels of awareness of industry developments, how the operation is currently managed, and what expectations are held for the future. The overall picture was one of optimism for the future, tempered by the initial caution that characterises any new field.*

## Introduction

The use of the World Wide Web (hereafter the web) by regional newspapers has already received some attention[1-6] but much of the available literature comes from the USA, where the regional newspapers cover a far larger geographic area and have notably more capital at their disposal, and where telecommunication tariff structures have resulted in patterns of web usage that are radically different from those prevailing in much of the rest of the world. Here, we provide an overview of the current status of the use of the web by regional newspapers in the UK, thus complementing two recent detailed studies of one such regional newspaper, specifically the web version of the *Northern Echo*[7,8].

A Yahoo! search identified about 40 different regional newspaper web sites: the 25 that are discussed here are all published on a daily basis and have a sufficient readership to enable them to consider launching a fully functional online version. The URLs of these sites are listed in Table 1, together with the parent regional newspaper by which they will be referred to throughout this paper. Each site was evaluated, taking account of the criteria identified by Bayne[9] for the evaluation of commercial web sites (such as content quality, ease of navigation, features, graphics, interactivity and update frequency, of which content quality and update frequency are clearly of particular importance in the present context). The sites were evaluated from the viewpoints of both the home and the business user, with all the sites being viewed with both Netscape Navigator *via* a networked PC in the University of Sheffield and Microsoft Internet Explorer *via* AOL on a home computer; following these evaluations, further data was obtained from the responses to a questionnaire survey. The reader is referred to the MSc dissertation by Neilson[10] for fuller details of the study on which this paper is based.

## Site evaluation: the reader perspective

### *Image and layout*

The overall appearance of the sites varies greatly. The *Bristol Evening Post* and *Western Daily Press* both provide examples of a traditional newspaper format converted online: for example, the local news headlines and accompanying photographs are situated on the home page, with links to other areas of the site such as classifieds and weather taking a secondary role. This approach is in contrast to that adopted for instance by the *West Midlands Express and Star* and the *Dundee Courier*, where the news section of the site appears to be accorded only equal emphasis with other sections such as Entertainment or Local Products and Services.

It has been noted[3] that the brand name of a newspaper may be one of its major advantages in making its move online; however, many of the sites evaluated are not named after their host newspaper. Examples of these include *www.westmidlands.com* (*West Midlands Express and Star*), *www.go2coventry.co.uk* (*Coventry Evening Telegraph*), *www.sheffweb.co.uk* (*Sheffield Star*) and the *www.thisis....co.uk* model favoured by many of the newspapers owned by the Daily Mail & General Trust plc (e.g., the *www.thisisgloucestershire.co.uk* site hosted by the *Gloucestershire Echo*). The contributing newspaper is generally given some prominence throughout a site through the inclusion of its name and logo on the home page and news sections, although it is possible that a visitor surfing the site could fail to realise which body was responsible for the provision of the local information and services listed.

### *Content quality*

Quality of content is of paramount importance in attracting and retaining visitors to a site but it is difficult to evaluate the quality of the journalism of an online version of a regional newspaper without being an established reader of the existing print edition. It was, however, noticeable that few of the sites list reporters' names alongside their articles: this is generally a formality in print newspapers, and its inclusion might go some way in helping to create an impression of integrity and accountability for the new medium. At a more general level, the sites showed few examples of errors such as spelling mistakes or outdated links, although it was noted that some still have rather a large proportion of their site under construction, without any expected dates for completion.

In theory, the online medium should allow almost instantaneous updating in a manner not possible in print. However, this is frequently not the case, with updates generally occurring only once daily (i.e. often less than the equivalent print publication) and with some sites providing no information to the reader regarding the time of the last or next update. The sites that provide the best information on updating include the *Dundee Courier* and the *Bristol Evening Post*, where it is clearly stated that the sites are updated at 16.00

*Table 1. The 25 regional newspapers that were evaluated during the study. (The nine starred regional newspapers were those that responded to the questionnaire.)*

| | |
|---|---|
| *Belfast Telegraph* | *http://www.belfasttelegraph.co.uk* |
| *Bolton Evening News* | *http://thisislancashire.co.uk* |
| *Bristol Evening Post* (*) | *http://www.epost.co.uk* |
| *Cambridge Evening News* | *http://www.cambridge-news.co.uk* |
| *Coventry Evening Telegraph* | *http://www.go2coventry.co.uk* |
| *Dundee Courier* (*) | *http://www.dcthomson.co.uk/courier* |
| *East Anglian Daily Times* | *http://ecn.co.uk* |
| *Essex Evening Gazette* | *http://essex-news.co.uk* |
| *Gloucestershire Echo* (*) | *http://www.thisisgloucestershire.co.uk* |
| *Grimsby Evening Telegraph* (*) | *http://www.grimsby-online.co.uk* |
| *Huddersfield Daily Examiner* | *http://www.examiner.co.uk* |
| *Hull Daily Mail* (*) | *http://www.hulldailymail.co.uk* |
| *Lancashire Evening Post* (*) | *http://www.lep.co.uk* |
| *Leicester Mercury* (*) | *http://www.thisisleicestershire.co.uk* |
| *Liverpool Echo* (*) | *http://www.liverpool.com* |
| *Oxford Mail* | *http://oxford.newsquest.co.uk/mail* |
| *Portsmouth News* | *http://www.thenews.co.uk* |
| *Sheffield Star* (*) | *http://www.sheffweb.co.uk* |
| *Staffordshire Sentinel* | *http://www.thisisstaffordshire.co.uk* |
| *Swindon Evening Advertiser* | *http://www.adver.co.uk* |
| *Torquay Herald Express* | *http://www.torquay-online.co.uk* |
| *West Midlands Express & Star* | *http://www.westmidlands.com* |
| *Western Daily Press* | *http://www.westpress.co.uk* |
| *Worcester Evening News* | *http://www.newsquestmidlands.co.uk/wen* |
| *Yorkshire Post* | *http://www.ypn.co.uk* |

and 18.00 hours, respectively, and the *Essex Evening Gazette*, which gives the time of the last update for individual news stories.

A comparison of the printed and online versions of some of the papers suggests that not all of the hardcopy stories appear online, and even when this is not the case, a fair degree of editing may be involved, e.g., the news stories on sites such as that for the *Liverpool Echo* or the *Leicester Mercury* appear to be summaries of the print versions. However, some sites provide other features that may compensate for the limited provision of current news stories. For example, *Torquay Online*, the *Hull Daily Mail* and the *Swindon Advertiser* include searchable archives of older news stories, a feature that will particularly appeal to local historians as the archive grows in size and becomes a valuable record of local information.

The increased space available in an online newspaper often results in a substantial expansion in the size of the section dealing with Events, Leisure and What's On guides. Thus, many of the sites studied have detailed directories of museums, pubs, clubs and restaurants that can be searched or browsed through: a good example is the comprehensive *Sheffield Star* site, where the leisure section includes items such as wildlife walks and special discount holiday offers for local readers. Local sporting activities are also well served online: some sites, indeed, seem to give almost equal prominence to sporting news as they do to the main news features. Football attracts special attention: the *Bristol Evening Post* has a link to the *Green 'Un,* an online publication devoted to local sports results, and a number of newspapers have a whole area of the site devoted to the local team, such as the Grimsby Town FC section hosted by the *Grimsby Evening Telegraph.*

Many other types of local information were found in the sites that were visited. Thus, the majority provide weather forecasts, with the *Belfast Telegraph* providing one of the most comprehensive with a five-day outlook, and the *www.thisis....co.uk* and *Liverpool Echo* sites also providing information on traffic, late night chemists in the area, local government links, help lines and emergency numbers. Other good examples include: the *Portsmouth News* and *Huddersfield Examiner* sites, both of which provide performance tables detailing the waiting lists for different specialities in local hospitals with the *Examiner* also publishing the OFSTED reports on schools in the locality; and the *Lancashire Evening Post* site, which supplies information on local council tax bands for property in the area. The provision of this sort of community information enhances the traditional role of the newspaper as information provider, allowing them to complement more traditional information disseminators, such as council departments and local libraries: indeed, the *Liverpool Echo* site describes itself as 'the gateway to the pool of life'.

### *Catering for a wider audience*

A web site provides an opportunity to attract new readers who might not take the printed newspaper, and it is clear that many of the sites are targeting two such groups in particular: ex-patriots and young children.

Readership of the printed edition of a regional newspaper can be difficult for people who do not live within the locality but an online edition is readily accessible to a global audience. Whilst UK regional newspaper sites may be of only limited interest to Internet users in general, they are of considerable interest (potential or actual) to expatriate readers who can now easily keep in touch with the news from their home town. Evidence of this interest from abroad becomes apparent from studying the discussion and feedback areas that are available on some sites: indeed, expatriates provide a large majority of the correspondents in such areas. Typical comments include 'I do get home every year but it's great to have 'on-line' news ... though it makes me homesick I still read it everyday and love to see what's going on my home town' (in the *Liverpool Echo* Guest Book) and 'I log on from here in Australia at least 3 times a week and find it a great way to keep up with all the news from Lancashire and around the country' (in the *Lancashire Evening Post* Forum).

In the online medium, sections of the newspaper with particular appeal for expatriates can be expanded and given a higher profile than in their print counterparts. For example, the *Dundee Courier* hosts the 'From Dundee and PROUD of it' campaign, the welcome page of which includes a message from the Provost of Dundee inviting people worldwide to become involved, and this site also links to clan pages, giving details of name origins and associated tartans. The *Liverpool Echo* site dedicates an area to its expatriate readership that has been 'created specially for expat Scousers', as does the *Belfast Telegraph*, which is probably one of the most comprehensive sites for expatriates ('you're never far from home') and has over 700 members listed. Its news columns provide stories from fellow expatriates about their experiences since emigrating (an idea also used on the *Bristol Evening Post* site), and it is possible to search their members' pages *via* a world map to locate fellow exiles. Visitors are encouraged to register, and can receive special offers such as discount deals at Belfast hotels. The Exiles Club also offers a subscription service to a weekly comprehensive news briefing, costing £30 annually, that can be ordered securely online. This provides one of the few examples of a chargeable service on any of the sites evaluated, but details of how many people actually take up this offer are not available.

Children are often more computer literate than their parents, and several of the sites have made special attempts to attract younger visitors who are not part of the traditional print readership. Thus, the *East Anglian Daily Times* site has a Kids section containing both educational (e.g., articles on the History of Rome, and Street Safety) and recreational (e.g., a 'Your Pictures' section that offers the opportunity for children to send in their drawings for online display) material. This mix is also apparent in the 'adult-free zone' of the *Coventry Evening Telegraph* site. This zone contains reviews of, and links to, other sites of interest to younger readers (largely those with content dominated by comics and computer games), an online discussion group, a resident agony aunt (CyberSteph), information on prospective job paths and CV advice, and links to outside web sites that may help with younger children's homework assignments. An alternative strategy is exemplified by the *Belfast Telegraph*, which tries to attract a younger audience through schools rather than on an individual level: the Schools Connection section hosts primary and secondary school web sites, the content of which is largely designed and provided by the pupils, and the Schools Discussion Area provides a forum for teaching staff and pupils to make enquiries and to air their opinions.

It is clear from examples such as these that many of the regional newspapers recognise the potential of this new audience. However, few of the sites provide the sophisticated interactive facilities that are likely to appeal to children. This may reflect the expertise and expense necessary to design such features, an important consideration in the context of a site that is providing free content.

### *Use of new technology and interactivity*

Exploiting the interactive potential of the Internet is seen as an important aspect for commercial sites, and regional newspapers must use this capability if their online version is to offer anything radically different from the print copy. At the most basic level, it allows readers to respond to what has been printed online through emails that are then posted on the site. The *Worcester Evening News* has a Letters & Opinions page for responses to editorial comments, and the *Portsmouth News* has a Feedback section for contacting editorial, sports or the newsdesk. Online ordering of various goods and services is provided by some of the sites, e.g. the *Liverpool Echo* and *Gloucestershire Echo* sites have sections for the sale of photographs and books, respectively, the *Belfast Telegraph* sells a Belfast 1998 calendar online, and it is possible to subscribe to the hard copy edition of the *East Anglian Daily Times* via its web site.

Email updates are another service offered by some online newspapers. The *East Anglian Daily Times* offers a free service for job hunters, whereby readers can opt to have updated vacancy listings in the relevant selected category sent to them. Sports fans who read the *Huddersfield Examiner* online can register for Sports News *via* email, and can cast their vote in polls on current sporting issues. The *Belfast Telegraph* offers the most sophisticated option, with its 'Push Channel' enabling the delivery of the latest news direct to the reader's desktop. However, in order to access this service it is necessary to download Backweb client software, and although the relevant link is provided, the download itself can be very time-consuming and take up a fair amount of memory space. This could mean that some readers owning older PCs might be unable to take advantage of this service, and similar comments apply to the multimedia video reports that are available to readers of the *Hull Daily Mail* who have downloaded the RealPlayer viewing software. However, such sites demonstrate an awareness of the greater role these applications are likely to play in the future as more sophisticated machines become widely available.

Personal classified advertisements provide an obvious area for interaction. The *Portsmouth Evening News, Lancashire Evening Post, Yorkshire Evening Post* and *Liverpool Echo* sites all have links to PersonalPages or the Personals Online Network, both of which are international companies hosting personal advertisements from around the world. These services enable readers to send confidential email, listen to other members' voice greetings online using RealAudio, and see suggested matches based on their own detailed profile.

Probably the most novel use of the interactive possibilities of the Internet is the creation of the Virtual Pub, as exemplified by the 'Mat & Mouse' at the *Coventry Evening Telegraph* site and the 'Crown Liquor Saloon' hosted by the *Belfast Telegraph.* At the Mat & Mouse, the reader can select which area of the pub to visit by clicking on the relevant handpull at the bar, the graphic design of which is excellent. Meet the Locals features famous people from the locality, the Pub Survival Guide details stories, jokes and tricks for performance in its non-virtual equivalent, Talk Back is a discussion list for regulars, and there is also a Pub Quiz (although its need for Java-enabled technology might limit rapid uptake). The Crown Liquor Saloon offers a feature called the Mug Punter, where readers can join in online betting on horse racing, once they have given themselves appropriate online aliases. The pub also has a live video link from the Crown Bar in Belfast, and in the Meet the Team section, a live webcam of the Internet team at work in the office; another example of a webcam is provided by the *Sheffield Star* site, which has a link to help commuters gauge the level of traffic congestion on one of the main roads into the city.

### *Ease of navigation*

Many of the sites studied have incorporated frames as a navigational tool, with the main subsections of the site listed either above or alongside the page being viewed in the central frame. In general, the sites were easily navigated and appeared well-planned, although there were some instances in which readability could be improved. The most novel manner of displaying the news is that of the *Yorkshire Evening Post*, where a miniature copy of the regular print newspaper is shown online. The paper appears to the left hand side of the screen, and pages can be turned by clicking arrows beneath the image. The full details of a story can be brought up on the right hand side of the screen by clicking on the relevant article in the scanned-in newspaper. A number of the sites (e.g., the *Dundee Courier* and the *Belfast Telegraph*) also offer a keyword search service, in order to ease quick access to the area of interest to the reader.

### *Monitoring traffic*

There was generally little evidence of visible counters on the sites evaluated to log the number of visitors although the volume of traffic may still be being monitored invisibly. However, a number of the sites are making attempts to compile a reader profile through registration, special offers and competitions. Thus, the *Dundee Courier* site runs a banner on its home page offering a free Dennis the Menace fridge magnet to its readers. When a reader clicks on the banner, they are asked to fill out a form detailing information such as age, location, occupation and salary in order to receive their free gift. The *www.thisis....co.uk* sites offer a number of services that are available only to registered readers. After filling in a form listing their personal details, visitors are allowed access to chat rooms and bulletin boards. Readers are also offered an opportunity to personalise the information provided by the site by filling in a hobbies/interests questionnaire that will enable prioritisation of items considered to be of special interest to them.

## Site evaluation: the business perspective

Some of the features listed below such as classified advertisements and business news are of obvious interest to both commercial and non-commercial visitors, and it is recognised that the sites evaluated were not split into two clearly defined areas. However, for ease of analysis it is helpful to make this somewhat artificial division in order to concentrate on the presence of commercial activity in online newspapers.

### *Business news*

The majority of the sites evaluated have a business news section similar to that provided for general and sport news. The volume of local business news stories may not be high enough to warrant daily updating, so older stories tend to stay online longer than other types of news story; it is perhaps for this reason that sites such as the *Western Daily Press* simply entitle their business news section Business Weekly in order to avoid confusion. The space and consideration given to business sections within online newspapers does appear to vary greatly, with some sites such as the *Dundee Courier* and the *Coventry Evening Telegraph* sites not providing any business news at all, as opposed to the high level of content on the *East Anglian Daily Times* and the *Bolton Evening News* sites.

### *Surveys, reports and links*

Many online newspapers publish a variety of additional content for their business audience. Commentaries on the state of industry in the region are popular, with the West of England Business Guide in the *Western Daily Press* featuring the Top 100 companies in the region and similar statistics. The *Essex Evening Gazette* and the *Bolton Evening News* have homepage links to Essex 2000 and East Lancashire 2000, respectively, organisations that aim to link business and enterprise in their regions and that provide detailed quarterly updates on developments in the area.

Special topic supplements and articles are also provided by many sites, especially in the area of new IT issues. For example, the *Bristol Evening Post* business section includes an IT and the Internet supplement, providing information on how local companies can best utilise changing technology, and the CourierNet section of the *Dundee Courier* online features articles on issues such as network PCs, teleworking, ISDN and online commerce.

As well as company and industry news, local businesses may need to access the latest financial and market news from the City. The *Western Daily Press, Portsmouth News, Cambridgeshire Evening News* and *West Midlands Express & Star* sites all provide regularly updated information on exchange rates and the closing prices of local and popular companies.

Useful links to outside organisations are provided by the *Bolton Evening News,* which lists contact details for local Business Link offices, Chambers of Commerce, Technology Enterprise Colleges and Technology Management Centres. Business Diaries of upcoming industry events are kept at the *Portsmouth News* and *Huddersfield Examiner*, and a digest of public sector contracts available in East Anglia is provided at the *East Anglian Daily Times* site. The provision of these types of features is helpful in creating a local information gateway, and therefore increases interest in the online medium from businesses in the region.

### *Classified advertisements*

The majority of the sites studied link to ADHunter (at *http://www.adhunter.co.uk/*), which is a database set up by the regional newspaper industry to provide a centralised search service for classified advertisements. A number of sites also provide local property advertisements; for example, the Prestigious Properties section of the *Dundee Courier* provides an interactive guide to local properties for sale, with access to photographs and floor plans. House-hunting over the Internet is growing increasingly popular with the amount of media attention it has attracted, and features such as these should enable the regional newspapers to overcome the threat of estate agents setting up their own interactive classified sites.

### Business directories and databases

Most regional online newspapers have set up a listing of local companies, searchable by company name, category and location. The *Bristol Evening Post* and *Western Daily Press* sites host a company database and also an email database of the contact details for business individuals. One of the most sophisticated business directory designs is found on the *www.thisis...co.uk sites*. After locating a business through name and sector searching, the reader is given details of the web site address, Internet fax and email where possible, and an interactive map provided in association with Ordnance Survey facilitates finding the geographical location of a business.

The majority of sites would appear to provide listing services such as these free of charge to local businesses, although it was noticeable that some of the sites seemed to contain only businesses that had already established a web presence. One exception is the Business Connections directory on the *East Anglian Daily Times* site, where firms have to pay from £50 *per annum* to register their web page on this resource and where over 100 companies are already registered.

### Advertising online

Many of the online newspaper sites studied have tried to attract advertising revenue, using a number of different approaches. One option is the provision of enhanced listings in company directories, a service that is offered on the *Sheffield Star* site. This enables companies to add a small sales or information message, possibly with some graphics content, to their basic listing, thus providing a small advert for the company, and creating a more eye-catching appearance than that of competitor links. The rates for this service are not provided online, but full contact details are listed. A different approach adopted by sites such as those for the *Torquay Herald Express*, the *Staffordshire Sentinel* and the *West Midlands Express & Star*, is to have a section of the site devoted to their advertisers in the form of an index, which provides descriptions of the advertising firms and links to their external web sites. An alternative format, as exemplified by the *Hull Daily Mail*, *West Midlands Express & Star* and *www.thisis...co.uk* sites, is to have a banner advertisement positioned across the top or bottom of a page, with a live link to an advertiser's home page.

An important aspect in attracting potential advertisers online would appear to be the active promotion of the possible opportunities for local businesses. A number of the sites provide detailed statistics regarding their print circulation and print advertising rates, but do not go on to promote their online version as an advertising vehicle. The provision of an online advertising rate card would appear to be a sensible method of informing interested parties, and these are made available on some of the sites evaluated. Figures quoted for an animated banner advertisement ranged from £7 per week to £150 per month, although one would expect that such price variations are likely to lessen as the web becomes better established as an advertising medium. Of more importance at this stage is raising the profile of online advertising, and many sites have a page dedicated to outlining the possibilities for online advertising and providing contact information.

A further promotional tool employed by some sites is to utilise the banner space itself to advance the possibility of advertising there, a technique already familiar on billboard advertising. When a banner site has not been hired by an external advertiser, the newspaper uses it to post a message similar to 'this space could be advertising your business', thereby alerting potential advertisers to the possibility of advertising online. Examples here include the *Grimsby Evening Telegraph*, *Torquay Herald Express* and the various *www.thisis...co.uk* sites, which link their promotional banner to the rate card page.

The provision of web services for local businesses provides a further source of income. Thus, a live advertisement on the home page of the *Western Daily Press* site links to a page detailing the advantages of getting your business on the Internet through the newspaper, with benefits such as high traffic, secure facilities for online commerce and search engine registration services. Sites such as that for the *Gloucestershire Echo* offer further detail and promotion of their services, with a Gallery of web sites designed and hosted by the newspaper, and quotes from various satisfied customers with links to their sites.

## Questionnaire responses

The site evaluations were complemented by a self-completed questionnaire consisting of 18 open-ended questions. A telephone enquiry was first made to ascertain the most suitable member of staff of each regional newspaper for receipt of the questionnaire, and their preferred method of delivery. In total, 20 email and five postal questionnaires were sent out for completion. Unfortunately, despite two follow-up reminders being sent, only nine completed responses were obtained and the conclusions that can be drawn as to the current state of the art can thus only be indicative in nature, with responses being obtained from: the *Bristol Evening Post*, the *Dundee Courier*, the *Gloucestershire Echo*, the *Grimsby Evening Telegraph*, the *Hull Daily Mail*, the *Lancashire Evening Post*, the *Leicester Mercury*, the *Liverpool Echo* and the *Sheffield Star*.

### *When was your online site launched?*

Responses here varied quite considerably, from August 1995 through to July 1998. One might have expected that the more established sites would be further developed, but the site evaluation would seem to indicate no obvious link between the sophistication of a site and its launch date, so there may have been some advantages in waiting for the medium to develop before investing resources.

### *How many people work within your Internet or New Media department, and how does this compare with the print edition?*

The resources provided to an online regional newspaper show considerable variation: one of the regional newspapers has no full-time member of staff on its Internet team, while another has no less than 12. These numbers are very low when compared with the hundreds of people normally involved in the production of a printed version; however, the questionnaire responses indicate a move towards expansion, with one of the responding regional newspapers planning a staffing increase from one member to seven in the near future.

### *Have you formed any alliances with local firms or organisations in this venture? Please give details.*

Commentators such as Auman[1] have mentioned the possibility of forming alliances with other organisations in order to help gain control of the market, and a number of the respondents indicate that they have followed this path. Thus, as noted previously, many of the regional newspapers are part of the ADHunter classified advertising operation, and other local and industry partnerships have also been formed. Examples include: membership of a Regional Independent Media group, with links to regional newspapers – and their web sites – in the locality; strong links with other publications owned by the same ultimate publishing house; local ISPs, sometimes with tariff reductions for readers; the local county council and the local university. Some regional newspapers had no alliances with any external organisations, one respondent noting that 'we are working with local organisations to commission development work on areas where we lack the necessary skills but I would not term these alliances. It is our intention to retain full control over any web site operated under one of our newspaper titles'.

### *Who is responsible for the design and maintenance of your site? How/why was this decided?*

The respondents were unanimous in declaring a preference for in-house control, with the only 'outside' influence being the head office of the ultimate publishing company in one case (and even here the local newspaper team has 'ultimate say' concerning site content). Some respondents had considered placing some responsibility in outside hands, but had rejected the idea for various reasons. For example, one respondent commented that 'A slow development by an outside agency which may not understand the pressures of news gathering would clash with our need to hit deadlines. We never thought we were in the business of solely creating 'eye-candy'!' and another one commented that 'This has been entirely handled in-house.... We have since invited external designers to submit fresh ideas but so far these have proved to be disappointing. We are currently initiating our own refreshment of the site'. The replies also indicate that responsibility for maintenance of the site often falls to only one or two people. This might be an existing member of staff with the necessary knowledge ('Our Editorial Designer...had developed his web skills over the previous couple of years and he had done some web design work on a freelance basis in his own time') or a new employee who had been taken on to carry out the operation ('They had to employ someone as Systems did not have the time to update and could not design ads. They contacted the local university to get me after graduation').

***Who do you regard as your main competitors in the provision of online news – now and in the future?***

The view that, 'we are aware new technologies may allow people to be a rival to our efforts electronically, both now and in the future' was shared by the majority of the respondents, although one asserted that, 'we really feel that we have no competitors as such, the local news industry is a new market to the Net, and we were the first'. Within the majority that do anticipate the existence or possibility of competition, the perceived rivals take various forms, with the overall consensus being that existing print competitors are the major current rivals online. Future competitors mentioned included: major players such as Microsoft, BT and News Corp; search engines such as Yahoo! (although one respondent felt that, 'the recent development of 'gateways' where search engines like Yahoo! feed news to users may not go local enough to be a direct threat'); business directories; and specialised property, motor and jobs sites. Generally, local newspapers appear to be fairly well informed regarding possible market developments and some are already discussing plans to protect their existing position: for example, 'We are considering making limited local news feeds available to other web site operators, on the basis that this would prevent the threat of new entrants into the news-provision market place.'

***How often is the news content on your web site updated? How does this compare to the print edition?***

The observations made previously in the site-evaluation sections are generally confirmed by the questionnaire responses. Three respondents reported that their online versions are updated at the same rate as their print editions, but the remainder gave a variety of replies, for example, only Monday to Friday online updating (when the print edition is published six days a week) and local news updates once daily (when the print edition might have as many as five editions in the same time).

***Does the news content of your online site differ from that of the print edition, e.g. edited or expanded format, use of visual images?***

Only two respondents stated that the news content of the online and print versions were identical, with the others finding it necessary to modify their news content in some way for the different medium. Thus four of them select a proportion of suitable news stories for inclusion on the web site, although the text of these stories does remain largely unaltered, while two others cover all of the stories, but in a highly edited summary format. Several respondents made the point that although the news content may be reduced in some way, the web site also offers lots of other features that the print newspaper does not, such as colour photographs, archive facilities and additional content (examples of which have been discussed previously in the site evaluations).

***Which particular features within your site do you consider most successful?***

The responses here mirror closely the site evaluations, with the following considered to be particularly successful: sports sections, particularly those concentrating on local football clubs; classified advertisements; areas of the site that allowed interactive input for readers; and areas that are targeted at overseas readers.

### *Are there any content areas which you intend to expand or improve in the future?*

Many answers here were similar to that of one respondent who replied, 'about 101! The possibilities are endless but again resources are currently very limited'. The sections cited by the various respondents include archiving, search facilities, classified advertising, local events, reader offers, sports and news provision. It was recognised that some parts of the sites needed development to reach their full potential, and that continual improvement was vital. 'Information is important but out-of-date or non-updated information reflects badly.' One suggestion put forward was the possibility of subscriber access, despite this being a model that some commentators feel has little chance of success[6]. A more commonly voiced plan was to incorporate 'higher levels of interaction and new technologies', with one regional newspaper having plans for interactive TV listings and puzzles, and an online shop. Another respondent also mentioned an eagerness to develop electronic commerce, but concluded that 'until the public's strong resistance to the fear of hacking is overcome, or the web industry comes up with a solution to encrypting credit card details easily, we may have to wait a while yet'.

### *How important is the interactivity of the Internet to your web site? How do you exploit this capability?*

The general consensus concerning inter-activity is that it is an extremely important area that is currently under-developed, but will be more fully utilised in the future. Current interactive features include readers' fora, classified job, car and property advertisements, chat areas, video images, personals, guestbooks, news polls, bulletin boards and instant feedback. Plans to increase interactivity include the imminent launch of an online shop featuring 'several hundred products from six initial retail 'guinea-pigs' ' but one respondent commented that 'our main IT support here is quite properly dedicated to producing newspapers and maintaining the systems needed, so such things as interactivity tend to take a secondary place at the moment'.

### *What level of traffic visits your site, and how do you monitor this? Was this level expected, and are you satisfied with it?*

Four of the respondents were unable to provide detailed information regarding the level of traffic visiting their site as this was considered to be commercially sensitive information, although they did feel able to say 'we have a satisfactory level' and 'we are satisfied with its increasing numbers'. Many of the sites had no real targets or anticipations regarding traffic levels: thus, 'we had no real expectations a year ago' and 'we had no idea what to expect, so are pleased', comments that are hardly surprising given the newness of the medium but that are likely to change rapidly as the technology matures.

Traffic statistics that were reported ranged from 20,000 to 60,000 'hits' per week. Readership appears to be increasing at present, with no falling levels recorded as yet, although one respondent highlighted the fact that attempting to increase traffic levels will ultimately mean more money being spent on the site: 'These things tend to reach plateaux: to climb to the next level we will need to invest more. The question then is how many more hits for how much?'

### *Do you attempt to monitor what type of reader your online site is attracting? What have you found out as a result, and how have you acted on this information?*

Quite a number of the sites do not at present carry out any systematic monitoring of their readership ('there is no great feedback and we have conducted as yet no survey into who, what or where our readers are' and 'there is no requirement to register so we have little indication as to who our online readers are') although several have plans to institute monitoring in the future. Reservations concerning the validity of audience monitoring exercises were voiced by a couple of the respondents, with one in particular appearing to have found the experience of little use, describing it as, 'too disturbing for words and a very dangerous area. You cannot use science to edit a paper, only instinct. Don't work for anyone who tells you otherwise and beware of marketing people'.

***Has the provision of an online edition of your newspaper affected print sales, or do you expect it to in the future?***

The response to the first part of this question was comprehensively in the negative. Some viewed it as having no effect at all on print circulation, and others regarded it as complementary to the traditional paper ('we see the web edition as encouraging the sale of the print product as we can use it as a promotional tool' and 'my own feeling is that it should enhance the attraction to the paper. Our web site should promote sales ultimately if it shows the paper in a good way'). Another reason for not being overly concerned about the effect on print sales was 'I suspect a lot of people looking at the Internet don't buy the paper anyway – I mentioned students (not big regional paper readers) and ex-pats, and the sort of people who are increasingly relying on their PC for info.' A number of the respondents also mentioned the fact that it is difficult to gauge the effect of online advances when the print newspaper market is undergoing change and often decline due to a number of different influences: 'The issue of newspaper sales in a time of changing lifestyles is a huge subject and is much discussed by papers throughout the world, irrespective of the web effect.'

***What online services do you provide for local businesses, e.g., directory listings, site hosting/design, advertising opportunities? How are these promoted?***

Some or all of the services noted in the question were available through most of the sites, and were promoted by a number of different means. The first avenue for promotion was publicity raised through the print publication, although limits on the space available had to be considered. Nearly all of the respondents cited their existing sales force as an important medium for promotion of online opportunities, 'though it takes time to train and educate them'. Another respondent mentioned the use of monthly training seminars to raise understanding and awareness of business on the Internet. Interestingly, targeting existing customers through traditional channels in an attempt to convert them to electronic commerce appeared to be favoured over promotion online.

***How much do you charge for different types of advertisement on your site?***

Of the respondents who completed this part of the questionnaire, only one regional newspaper follows a policy of charging per hit, with a going rate of £30 per thousand page requests. All other respondents charge solely by inclusion, with further charge for any updates needed. A number of the sites have split rates according to the area of the site in which the advertisement is to appear, e.g., one regional newspaper charges £500 per year for the front page, £420 for news or sport, and £360 for any other area of the site. Many of the newspapers also offer the option of static or animated advertisements, with the latter costing between £50 and £100 extra. Some newspapers have yet to issue a standard rate card for advertising charges: for example, one respondent fixes prices on an individual basis, reflecting the need for initial flexibility in an uncertain market ('We have hesitated to publish a fixed rate card at this stage because we do not know enough about what our advertisers need. We evaluate a client's needs and discuss what we think is fair').

***How have your traditional print advertisers responded to the possibility of advertising online?***

The responses to this question varied markedly. Three stated categorically that the reaction had been 'positive', and another that 'it has been fully welcomed'; others, however, admitted that the response has been somewhat more complicated and mixed: 'Very mixed reaction, a few very enthusiastic, most apathetic, some don't know what we're talking about! Even those who say they're interested are difficult to close.' The level of interest varies according to the business sector of the company, with IT companies such as computer equipment suppliers responding positively, whereas other businesses such as estate agents have been far more cautious: one respondent noted that the latter 'are used to getting an immediate and definite response to their in-paper advertisements from a well-proven medium. They don't waste money, and still feel that it is best to leave involvement with the web for at least two years. (If you can

PROVE it works I'll have some!).' One of the respondents noted that, 'the big problem is the current gap between html structures and in-paper advertisement styles, and the transfer between the two'. The online advertisement does not contain the wealth of small detail that a traditional newspaper approach encourages, relying instead on the use of colour and animation. This can cause problems with advertisers who are reluctant to modify their approach: 'If the advertiser loves 'bitty' ads, it's hard to convert them. Education is the key here'.

***Is your site currently profitable? Do you expect it to become more commercially successful in the future?***

Of the six respondents who volunteered information here, two regional newspapers are running at a loss at present: however, neither seems overly concerned by this, and both have a time scale within which they intend to turn this around. The others are 'just about' profitable or 'definitely making a profit at the moment'. All the respondents appear to be confident that there will be a growth in levels of commercial success in the near future as a result of greater investment and new features that are planned, although one respondent commented that 'it's hard to plan beyond the next two years because of the effects of digital teletext and interactive broadcasting generally.'

***How do you attempt to increase awareness of your online edition? How is it advertised and through what medium?***

The most common means of promotion is through the print title, with all respondents utilising this possibility, and one regional newspaper relying solely on this method. Also popular is promotion on the Internet, through registration with search engines, placing links on other local pages, and promotion on the pages of sister publications. Other approaches mentioned included attracting people in the locality who do not read the print publication through placing bills in newsagents' windows, hosting roadshows, and attendance at appropriate exhibitions and trade fairs ('taking a live Internet feed with us to allow people to search the databases'). One respondent stated that direct mail promotion had been considered but 'the flaw here is that we do not know how many people access the Internet. Figures vary wildly'; this regional newspaper's hit rate suggests that most readership is midday and late afternoon, implying an office/school readership that is leading them to target their promotional activities at local schools and businesses, rather than at casual home use.

## Conclusions

The online versions of regional newspapers are all much newer, and are run on a much smaller scale, than their print versions. However, differences between the two media are already clear, with the online regional newspapers offering the possibility of attracting new audiences (e.g., ex-patriates and local schoolchildren) and improving both the content coverage (e.g., by providing searchable archives and hot links to associated sites) and the mode of presentation (e.g., by the interactive pub and facilities for property searching). Few of the sites are commercially successful as yet but there is an air of optimism about the future, especially if investment levels can be increased in what is currently only a very small part of the regional newspapers' operations. It will be interesting to see how the sites develop over the next few years.

## References

1. Auman, A. 'Not-so-strange alliances and their impact on online news media.' *http://www.december.com/cmc/mag/1997/auman.html* [visited 14 May 1998]

2. Bennett, R. 'Newspapers and the Internet.' *http://www.coams.com/ Outlook/BENNETT.htm* [visited 17 November 1998]

3. Coopers & Lybrand Media Group. 'Regional Press Report – the role of the regionals in the digital future.' *http://www.uk.coopers.com/coopers/media/ regional index.html* [visited 5 May 1998]

4. Goodin, D. 'Surfers flocking to local content.' *http://www.news.com/News/Item/0,4,14385,00.html* [visited 17 November 1998]

5. Mings, S. 'Online newspapers as familiar artefacts in new settings.' *http://www.december.com/*

*cmc/mag/1997/jul/mings.html* [visited 17 November 1998]

6. Outing, S. 'Newspapers on the Internet: advice for the future.' *http:// www.planetarynews.com/ sweden96.html* [visited 17 November 1998]

7. Williams, P. and Nicholas, D. 'The Internet, a regional newspaper and the provision of 'value added' information.' *Aslib Proceedings* 50, 1998, pp.255-263.

8. Chapman, P. 'Developing This is The North East.' *Aslib Proceedings* 50, 1998, pp.264-266.

9. Bayne, K. *The Internet Marketing Plan.* (Chichester: John Wiley & Sons, 1997.)

10. Neilson, K. *Regional Newspapers on the Internet.* (MSc dissertation, University of Sheffield, 1998.)

## Acknowledgements

We thank the respondents to the questionnaire, Bill Town for first suggesting this topic, David Miller for comments on an initial draft of the paper and the Humanities Research Board of the British Academy for funding Kara Neilson.

# The changing information environment in the media: case study *The Guardian/Observer*

*Helen Martin*

*Information Manager, Research and Information Unit, The Guardian/Observer, 119 Farringdon Road, London EC1. Email: Helen-Martin@Guardian.co.uk*

**Abstract** *Charts and evaluates changes to The Guardian/Observer library over the period 1983 – 1998. Examines these developments in the light of wider changes to the newspaper industry as a whole. Discusses the early arrival and impact of FT Profile; the setting up of the Association of UK Media Librarians; the increasing professionalism of newspaper library staff; computer assisted reporting (CAR), changes in service provision and a consequent change of name for the library, and the role of the Internet and intranet in the work of the unit. Identifies areas of future change.*

## Introduction

When asked to write an article about change in *The Guardian/Observer* Research and Information Unit over the last 15 years one understands full well what is expected; a detailed resume of technical changes over more than a decade; an academic assessment of their worth. But looking back it is not so much the technical changes that spring to mind, rather the changes in personnel and the qualifications required of them; the changes in the kind of information reporters want to retrieve and the depressing realisation that whatever the changes wrought in the immediate department – and they are considerable – the lack of parallel progress in the way our employers at management or editorial management level view us, is zero to negligible.

## Information Technology arrives

It was in 1983 that a company then called Datasolve (now FT Profile) invited all Fleet Street librarians to an office somewhere in Fleet Street to view their new product, of which they were very proud. To someone from *The Guardian* library, as it was in those days, it was an impressive toy. Always a small outfit, which had moved to London from its Manchester home in the early sixties, we had never been able to afford the blanket cutting that went on in other Fleet Street libraries, where armies of employees selected, cut up, folded and filed away literally thousands of newspaper cuttings every week. If a letter to the editor was signed by 20 people then a copy went into each of those people's biographical folder. Not so at *The Guardian*, where we took the two top names and hoped for the best. Now Datasolve had a system (World Reporter) where all 20 names were 'filed' automatically. If there were a few blips in those early days such as not being able to search South Africa, as a phrase, then that seemed, in those heady days, a small price to pay for the luxury of a full text searchable data host – even if it did not host many British newspapers at that stage.

The then librarians of *The Mirror* and *The Mail* were hugely unimpressed, maintaining – probably justifiably by the standards of their cuttings libraries and the size of their staff – that the system was too slow and cumbersome and that paper retrieval, astonishingly, was faster. Not even a little bit dissuaded, *The Guardian* librarian beat an excited path to the Managing Editor's door only to be met with a dismissive response. But attitudes were soon to change when it was realised that there was money in it for the paper as a provider of data. And early in 1984 the first World Reporter terminal was installed, not in the library, where it was feared that the staff, operating within a strict (union) closed shop, would become proprietorial about its positioning, but firstly in the Managing Editor's office and then, when lack of access became too much of a problem, in neutral territory in an unused office.

It was around this time too, that *The Guardian* became the first paper to appear on what was to become FT Profile. It was a seminal moment for the library, a rubicon, year zero.

All change dates from this moment. We had, in effect, our own internal electronic archive. It cost us, in those days, nothing, the cost being met by the royalty payments and the advantageous rates obtained for signing up first with Datasolve. The endless trawling through incomplete files was over. The hit and miss effect of our limited cutting exercise was laid to rest, even though we carried on cutting. The special requirements of newspaper libraries – speed and currency – were amply met by the system. The time taken to guess the right file (one of ten thousand possible headings), page through the cuttings, think of another one, page through that, find an article that answered the vital question – 'was the gun fired twice or three times'- meant that our failure rate had been previously high and the deadline often long gone by the time we had an answer. Now the answers were nearly instantaneous. Better still even NOT having an answer was instantaneous. We knew straight away if we had to pursue another avenue that did not involve newspaper articles.

The row about location of the online terminal rumbled on. As library staff became more familiar with the system, as journalists' queries began to be accompanied with a printout along with a file of cuttings, it became an irritation to have to go to another office to conduct the search. Protracted negotiations broke out about journalistic access to the system, an early precursor of the intermediary versus end-user row, even though there were no end-users at that time. In a typewriter-driven office that was not possible. But there was a row going on within the library too. The closed shop was beginning to break up all over Fleet Street as newspaper managements exploited the events at Wapping and the unions were nervous and edgy. World Reporter should not be brought into the library without extra pay and guarantees about job security argued some librarians, with perhaps understandable concerns, whilst to others amongst us it was obvious that if end-user use was adopted before we had managed to get our act together, the library would be isolated for good. Library management was of the opinion that it ought to put itself in the position of being able to embrace the new end-users, by training them and showing them the way. It sometimes seemed as if the enormity of the changes ahead had simply not registered with people and one can draw a parallel between that and the effect of the Internet today.

Eventually an arrangement was negotiated that gave the library a year's head start before end-user use became the norm. In the event, the delay of the installation of the editorial system Atex, extended that to three years, ample time to ensure professional proficiency. But attitudes had changed and maybe hardened as a result of the struggle and it became clear that in order to have a staff that fully understood the implications of the new systems we were working with, who could see the future and embrace it and who would even come to us fully cognisant of it, we would have to challenge the time-honoured procedure of Fleet Street recruitment, whereby vacant jobs were registered with the relevant union. There were few professional librarians on Fleet Street union books in those days but professional librarians, we became convinced, were what were needed if we were to survive. Post, rather than prior-entry, union membership would have to be encouraged. We began to advertise for staff elsewhere. The ease with which one records that decision defies the difficulty of accomplishing it.

Time, however, has proved our hunches right. The library increased in size, not decreased, with 'new technology'. With end-use of FT Profile now long established, around 81% of reporters still bring their searches to us at some time or another, in spite of desktop access. And by and large it has been those media libraries which embraced the end-user revolution and led it, which survived and prospered. It is hard to feel complacent however as a new generation of changes sweeps over us, but nevertheless extremely careful recruiting has been the bedrock of our policy ever since and one that has paid huge dividends – if not unfortunately to the talented staff. In the space of twelve years *The Guardian* library had moved from asking for no particular qualifications, to a degree in any subject, to a degree in librarianship, to an MA in librarianship/ information management, to, more likely today, an MSc in information science.

But World Reporter was to prove a catalyst in other ways too. Relationships between Fleet Street libraries had been cool and spasmodic up until then. Occasional phonecalls were exchanged if we needed things from each other, sometimes we had lunch with one another, but there was no structured link and the competitive nature of the relationship was always present. However, the enormous changes that all were undergoing, either by choice or imposition, or a bit of both, put pressure on us to share experiences and expertise and attempts were made to kick-start an old idea of forming a media library association.

The original idea had come from *The Financial Times* librarian and a few aborted attempts had been made in conjunction with the BBC and *The Guardian* – the latter two hosting preliminary meetings over lunch. But we lacked someone with enough time and real enthusiasm. These efforts dovetailed neatly into an approach to *The Guardian* library by Dr David Nicholas, then of the University of North London (UNL), to participate in a British Library funded study (Information seeking in an information society project) which, amongst other things, looked into the effect of the new technology on media libraries and journalists' information seeking habits[1,2]. *The Guardian* participated willingly, the first of many similar academic exercises that were to become the norm, and that were to add so much to our understanding not only of how our reporters worked, but also, and oddly, about how we ourselves worked. The act of recording the detail opened up a million doors and highlighted practices, some good, some bad, that were lost in the anonymity of daily routine.

## The Association of UK Media Librarians

The British Library investigators, who had also been researching television companies, offered to host a media library lunch at UNL, and present at that lunch was energetic American Sarah Adair, librarian at London Weekend Television (LWT). The original newspaper nucleus met subsequently, this time with Adair present. Soon we had linked up with a regional group led by Peter Chapman of *The Northern Echo*. In no time at all the Association of UK Media Librarians (AUKML) was born. The effect of the group cannot be underestimated. Let us not forget that professionalism in newspaper libraries in those days was not the norm. Newspaper librarians tended not to belong to Aslib or the Library Association and it might be fair to say, too, that even had they done so, their needs, then, were not well catered for by any existing association. The first meetings were heady with the novelty of having a forum in which to share our problems and attempt to solve them. Our needs were so specific – the vast amounts of data we held, the dependence on full-text, the uselessness of bibliographic databases, the currency, the speed at which we worked, the detail of retrieval. People, least of all database companies, rarely understood that it was not articles per se we searched for but the content of them – who had said what, exactly, when etc. Word proximity searching was one of our most important requirements. The oft-made comparison with hospital records was simply not relevant.

## Changes in service provision

Being data providers (to FT Profile and others) posed new problems internally – the preparation of the data feed was a new routine which today takes more time than the daily cutting process – roughly between 80-100 hours a week. To cope with this and allied to the decision to employ professional staff, we decided we would also start a training scheme and to this end employed two (now four) graduates on an annual basis on condition they went on to library school to study. The scheme proved to be beneficial, allowing us to offload the more routine tasks, (albeit we took care to vary the work) thereby freeing up our new qualified; staff to research. An added bonus which we had not really anticipated was that it also supplied us with an employment pool and at one stage 50% of the staff were former trainees; it made for great stability in staffing. The new-found time of the permanent staff, meanwhile, was put to good use. We had already begun to tailor-make services according to users' needs. Foreign correspondents were faxed a daily news service which told them of the major stories in their patch that day – their difficulty being, in pre-Internet days, in knowing whether their own story had made it into the paper that day and with what prominence, but more importantly in knowing what their colleagues on rival papers were doing. Many of them were too far away to buy copies of the paper on the same day.

On the editorial system we had introduced a library queue into which were placed mini-databases and chronologies on all kinds of current topics – IRA bombs being a popular and well-used subject. The idea was for reporters to have to hand quickly (and probably at a time when the library was not staffed) details of the last three bombs, say, without having to go online or search through files. For the most part we pre-empted their needs, but occasionally reporters would suggest subjects to us and occasionally, too, we would download the contents of a useful search. The criteria for the choice of subject matter was that it had to be something that would be used repeatedly. As the years passed the number of files on the system grew and the databases became comprehensive and extremely useful. It was not realised then that these databases would form the backbone of our editorial intranet.

### Computer assisted reporting

As AUKML expanded, along with our academic links, we established more and more contact with media librarians (and journalists) around the world. An American innovation that greatly interested was something called 'computer assisted reporting' (CAR). CAR is often a term employed by media people whose press is largely local and who have long been used to in-house electronic databases, to mean the searching of external data hosts. On *The Guardian* we had long been used to that and had added LEXIS-NEXIS, Context and Datastream to our array of online services. We, in fact, put the 'elitist' interpretation onto CAR – the building of relational databases on specific subjects in order to generate a story. The ease of doing this in the United States was one we envied. The access to official information, often in digital form, posed no problem for them. And the essentially local press provided a host of local issues to concentrate on (data on accidents, say, merged with hospital statistics and local transport figures, could highlight blackspots and solutions to them) which were far too parochial for a national press to bother with. But the principle was interesting.

There was no Freedom of Information Act in the UK, little was held digitally and the demands of a competitive national press were more taxing. A cautious approach to CAR was then taken. An enthusiastic member of staff undertook an Open University course in database construction and then proceeded onto Access. Starting with a specific breaking story, we began to collate the data on Access, hoping this would throw up links that had not been noticed. It did not, but we went on to expand the database into a parliamentary one, with election details and added on census information as well. Martin Linton, now MP for Battersea, but then a *Guardian* political reporter, used this information as the backbone of *The Guardian Election Guide* published by Fourth Estate. More recently our New Media Lab have expressed an interest in an adaptation of it for *The Guardian* web site.

Labour's promised introduction of a Freedom of Information Act had been eagerly anticipated and we were disappointed to be so badly let down, but the web opens up new possibilities for CAR and we continue to pursue it – the goal being to set up teams of librarians and investigative reporters.

### The Internet

Internet use at *The Guardian* started in 1992 although it took until the end of 1996 to get the coveted ISDN link. Once again it's introduction has been shadowed and followed by a British Library study (The Changing Information Environment[3]). Following a visit to the European School of Journalism in Maastricht by senior library personnel, for a week long seminar, led by the redoubtable Nora Paul, of the Poynter Institute, we revamped our inadequate training course, held three 'library lunches' by way of introduction to it and started some serious surfing.

At the beginning of the British Library project (April 1997) use was patchy, even in the library. We are recorded as saying we used it simply for quirky searches. By the end of 1998, however, there were clear signs of displacement on some of the external databases we use and also in hard copy and increasingly we use it not just for research and sometimes as a sole source (for song lyrics, say) but also for sub-editors' checking. For instance, 'does PolyGram have a capital G in the middle?' The answer on an external host like FT Profile may differ. The PolyGram site knows the answer, as of course do various books, but we find ourselves more and more reluctant to rise from our desks to use them! (Desk-top access to the Net is currently being rolled out on the back of our new editorial system, QPS, and by the millennium will be complete.)

## Intranet

Another place that knows the answer to the query is the hard-copy *The Guardian* style book. But books get lost and are hard to update without reprinting. We started building an editorial intranet in October 1996, using, as mentioned previously, information created on Atex, our editorial system. The databases and chronologies are now in part replaced by links to web sites in order to reduce maintenance. We will add such things as the style book, access to our new internal electronic archive (incorporating what Fleet Street call the Data Exchange – a free exchange of data between participating newspapers which is due to launch in 1999 using Verity); information on the Research and Information Unit and in time, canteen menus, local pubs and restaurants, personnel and other corporate information. A pan-company decision to adopt Lotus Notes means we will have to transfer our data from Front Page.

## Conclusion

So what does the future hold? The advent of *The Guardian/Observer* electronic archive has slowed down other projects for a while and raised the usual old chestnuts of does *The Guardian* need a library at all. It is hard for the men in suits to understand the intricacies of information provision. Hard for us to see why the partial replacement of one external host by another internal one should have any serious effect on the way R&I operates, or rather the use made of it, given we have had end-user searching for many years. There was plenty of time to anticipate this reaction, however, and in an effort not to be taken over by (the somewhat tedious inputting) archiving demands, come the day, we changed our name to Research and Information Unit, specifically so we could say 'Research' when answering the phone. It was hoped that this message might echo through editorial management's brains almost subliminally. Attempts were also renewed to have research the Unit conducted, where it stood alone, acknowledged and credited – an occasionally fierce fight. Newspaper journalistic ethics often manage to ignore blatant plagiarism! Television seems more generous.

Since the first real changes in 1984 when the library ceased to be a passive repository of cuttings and made tentative steps towards becoming a pro-active research orientated library, conscious efforts have been made not to be dictated to by technology but to harness it to our own ends. In some areas, like the adoption of external hosts, or CAR, we have led the way, in others like the creation of internal electronic archives we have lagged behind some newspaper colleagues. Above all we have tried to be user-based, tried to listen to our reporters and understand their jobs and needs and build a service based around those expectations.

The advent of *The Guardian* web site heralds more changes however. It is the web site that has driven the archiving project and in a fiercely competitive market newspapers look to all departments to make money. R&I has been making that money for 15 years now, editing the data we send to various hosts, providing data for books and now for the web sites too, almost as an aside. But whilst we must now turn our attention more to the commercial side of things we would be bitterly disappointed to lose our editorial roots and the expertise we have built up. The coming of the archive will free us from our need to cut up papers (purely as an issue of record these days) and more importantly free us from the confines and boundaries of R&I itself. We would like to see library researchers attached to projects and people and floating freely around the news floor; to see cyber-librarians on each 'desk' (foreign, home, features, sport) offering advice on web sites and anything else that is information based. We would like to develop our CAR teams and expand and maintain the intranet. Progress is slower now as these changes require editorial consultation but we edge towards them with occasionally clenched teeth.

Training is another issue that has taken on a new importance – training in external hosts, training in the Net, training on the intranet. The archive itself will need to be managed and can command a department of its own, particularly if we ever start to archive the web site.

News is changing. The broadsheets are dumbing down. Our intranet is full of 'kiss and tell' type chronologies that would not have been given house room 20 years ago. The 'Daily Me' is here and deliverable. Deadlines are vanishing to be replaced with 24-hour news. Editors are more at the whim of their readers than ever before and readers are becoming their own editors. It would be a fool who did not recognise the impact this is bound to have on newspapers and their libraries.

Ironically perhaps, however, because newspapers have always been a little dismissive of their libraries, a little reluctant to acknowledge a rogue professionalism in their midst, Media Labs have a growing appreciation of information skills. From the simple things like ordering their immense bookmark lists into thesaurus grouping, to the huge dependency on search engines, information management seems to have finally come home. This is high-tech stuff of course and may not be common-place for a few years yet, nor will it attract the librarian who likes the hands-on buzz of a news-room. But it will see out the dinosaurs and newspaper 'freaks' amongst us and usher in the millennium information manager.

We have travelled a long distance in the last 15 years – from paper to electronic databases and the Net, from no qualifications to the highest possible, from passive to proactive – and slayed a few dragons en route but the media librarian is not dead yet.

## References

1. Nicholas, D., Harris, K. and Erbach, G. *Online searching: its impact on information users.* (Mansell, 1987.)

2. Nicholas, D. *et al. End-users of online information systems: an analysis.* (Mansell, 1988.)

3. Nicholas, D. *et al. The media and the Internet.* (Aslib, 1998.)

## Subscriptions

*Aslib Proceedings* is published ten times per annum and supplied to corporate members of Aslib. For 1999 the annual subscription rates are £170/$280 (£136/$224 to members of Aslib). For subscriptions and back issues please contact: Portland Press Ltd, Commerce Way, Whitehall Industrial Estate, Colchester CO2 8HP, United Kingdom. Tel: +44 (0) 1206 796 351. Fax: +44 (0) 1206 799 331.

---

Aslib, The Association for Information Management, is a world class corporate membership organisation with over 2000 members in some 70 countries. Aslib actively promotes best practice in the management of information resources. It lobbies on all aspects of the management of, and legislation concerning, information at local, national and international levels. Aslib provides consultancy and information services, professional development training, conferences, specialist recruitment, Internet products, and publishes primary and secondary journals, conference proceedings, directories and monographs.

Further information can be obtained from: Aslib, The Association for Information Management, Staple Hall, Stone House Court, London EC3A 7PB, UK. Tel: +44 (0) 171 903 0000. Fax: +44 (0) 171 903 0011. Email: aslib@aslib.co.uk *http://www.aslib.co.uk/*

---

Periodicals Postage Paid at Rahway NJ and at additional mailing offices.
**POSTMASTER:** Send address changes to PROCEEDINGS, c/o Mercury International, 365 Blair Road, Avenel NJ 07001

---

©1999 Aslib and contributors ISSN 0001 253X

ISSN 0001 253X
Volume 51
Number 4
April 1999

21/2/2000
wcl

# Aslib Proceedings

## New information perspectives

### Information retrieval and the perpetual innovation economy

Murat Karamuftuoglu

### Internet, intranet, extranet: managing the information bazaar

Ian Watson

### A national survey of the citizenship information needs of the general public

Rita Marcella and Graeme Baxter

### The migration of news to the web

Peter Williams and David Nicholas

Aslib THE ASSOCIATION FOR INFORMATION MANAGEMENT

*Aslib Proceedings* is a long-standing professional journal of international repute, which has a large and wide subscriber base. The journal brings currency, authority and accessibility to the reporting of current research, issues and debates in the broad area of information work. Above all the journal wishes to provide research and comment in a form that is easily and quickly understood: a fresh, rigorous, but unfussy, writing style is what is aimed for. The needs of busy practitioners (and academics) is very much in mind. Articles are now refereed to meet the growing need for accredited and authoritative information. To ensure that this does not result in an unacceptable loss of currency, articles are speedily refereed by the journal's large, eminent and multi-disciplinary Editorial Board (assembled for just such a purpose). In exceptional cases they may be referred to an external panel of consultants. The journal's monthly publishing pattern helps in maintaining the currency of reporting. In consequence the submission of research in progress articles is encouraged.

The boundaries of information work are rapidly changing and widening. *Aslib Proceedings* has a role to play in the repositioning of the information profession by encouraging submissions from information managers, providers, packagers and consumers of all kinds – and especially those from the related information fields of journalism, electronic publishing, communication and Internet studies. An interdisciplinary Editorial Board has been assembled to assist in the process.

Prospective authors should submit articles of 2,000-6,000 words in length. Articles should be accompanied by a 100-150 word informative abstract and the numerical form of referencing should be adopted. In style and tone, submissions should fall between that of the journal's sister publications: *Managing Information* and *Journal of Documentation*. Articles should be furnished by email or on disc to David Nicholas, Editor, Aslib Proceedings, Department of Information Science, City University, Northampton Square, London EC1V 0HB. Email: nicky@is.city.ac.uk. The Editor is happy to provide advice on an informal basis to potential authors.

**Editorial Board**
Dr David Nicholas, *Head of Department, Department of Information Science, City University;* John Akeroyd, *Director, Learning Resources, South Bank University;* Sarah Blair, *Head of Publications, Aslib;* Peter Chapman, *Information Manager, NewQuest North East;* Professor Peter Cole, *Head of Department, School of Journalism, University of Central Lancashire;* Dr Tom Dobrowolski, *Senior Lecturer, Institute of Information Science and the Book, University of Warsaw;* Harry East, *Senior Research Fellow, Department of Information Science, City University;* Mike Hyman, *Managing Director, Head Software International;* Robin Hunt, *Creative Director, AreHause;* Nat Lievesley, *Network Manager, Centre for Policy on Ageing;* Helen Martin, *Information Manager, Research and Information Unit, The Guardian*; Dr Ian Rowlands, *Lecturer, Department of Information Science, City University;* Richard Withey, *Director, New Media, News International.*

**Editorial Team**
*Editor:* David Nicholas
*Deputy Editor:* Ian Rowlands
*Editorial Manager:* Kay Nicholas
*Editorial Assistants:* Peter Williams (City University), Colette Coles (City University)

# CONTENTS

99-108 Information retrieval and the perpetual innovation economy
Murat Karamuftuoglu

109-114 Internet, intranet, extranet: managing the information bazaar
Ian Watson

115-121 A national survey of the citizenship information needs of the general public
Rita Marcella and Graeme Baxter

122-134 The migration of news to the web
Peter Williams and David Nicholas

# Information retrieval and the perpetual innovation economy

*Murat Karamuftuoglu*
*Department of Information Science, City University, Northampton Square, London EC1V 0HB*
*Email: hmk@soi.city.ac.uk*

**Abstract** *The main objective of this article is to show the increasing relevance of the knowledge production capability of information storage and retrieval systems in the context of 'perpetual innovation', otherwise known as the 'information' economy. The knowledge production potential of information retrieval systems is only barely recognised in the information science community. Traditionally, information professionals and retrieval systems devised by them are conceived merely as guardians and facilitators of knowledge. This prevents information professionals playing a key role in an innovation based economy. In a perpetual innovation economy, information/knowledge embedded in commodities becomes the main source of profit. However, the peculiar character of information/knowledge means that privately owned knowledge tends to flow back into the public domain. This peculiarity necessitates continuous production of new knowledge applied to products and production techniques. Creative acts are not individualistic but collective/collaborative processes. Emerging collaborative systems on computer networks, such as the Internet, make it possible to devise work environments that are conducive to the development and cultivation of collective practices. Informational retrieval systems designers and practitioners may find it useful to study such systems to develop retrieval mechanisms that enhance creativity and facilitate knowledge production as well as knowledge transfer. It is hoped that by putting information retrieval in the context of the perpetual innovation economy, the knowledge production potential of information retrieval systems becomes more widely acknowledged and accepted among information practitioners.*

## Introduction

The objective of this article is to show the increasing relevance of the knowledge production function of information retrieval in knowledge based economies or, as they are sometimes called, innovation economies. Traditionally, the role of the information profession is seen to preserve and maintain knowledge contained in documents and other information objects and facilitate access to knowledge. From this perspective information retrieval is thought to be a process of transfer of knowledge stored in databases to the users of retrieval systems. This view overlooks the knowledge production potential of information storage and retrieval systems. It is the argument of this article that knowledge production is an important aspect of information retrieval, and it becomes particularly relevant in the context of the current developments in industrial production and economic relations. By drawing attention to current developments in industrial production and the role played by knowledge in economic activities, this article argues for the importance of conceiving information retrieval as a knowledge production as well as a transfer process. At present, the knowledge production capability of retrieval systems is not widely recognised in the information retrieval community or among the information practitioners in general. It is the overall objective of this paper to help facilitate a better appreciation of the importance of re-conceptualising information storage and retrieval from the perspective of knowledge production, by positioning it in the context of the emerging innovation or information based economies.

## Historical development of the capitalist mode of production: from extraction of surplus labour to technological unemployment crisis

There seems to exist a general consensus that knowledge has been replacing human labour and material resources as the driving force of economic development in industrialised societies for many years. Societies which are increasingly dependent on knowledge as a primary resource in their economic development are usually referred to by such labels as 'information', 'post-industrial', 'postmodern', or 'late capitalist' societies. The common claim implied by these labels is that, it is essentially the knowledge or information embedded in the industrial product that generates value and thus the main source of profit in such economies or societies. The defining characteristic of information economies is the high level deployment of information and communication technologies (ICT) in industrial production processes. In the US, for instance, investment in ICT has doubled in the last 10 years[1]. In Japan, use of numerically controlled machine tools, as a percentage of total number of machine tools, has increased from just under 50% in 1980 to over 75% in 1990. Similarly, production of robots in Japan increased from 19,873 in 1980 to 76,867 in 1991[2].

There are very different visions associated with the coming of the information society by different theorists, usually depending on the political alliances of the authors. For some theorists the information society heralds the beginning of an era where humans are finally liberated from performing routine, uninteresting and often dangerous jobs and human potential can now be channelled to more creative and intellectually demanding tasks. Conversely, others see in the increased mechanisation/computerisation of the production process the loss of the human labour capacity, increased levels of unemployment and alienation of the human subject.

A historical account of the development of capitalist societies can be found in the work of Marx. It is argued that this analysis is particularly relevant in understanding the development of the capitalist economies into what is usually called the 'information economy' (see the volume by Davis, Hirschl, and Stack)[3]. This section is intended as an overview of the ideas of such theorists as Morris-Suzuki, Carchedi, and Caffentzis who make use of insights acquired by Marx from his analysis of the production process in nineteenth century Western Europe to examine the development of production relations in information societies.

According to Marx, the generation of profits in capitalist economies depends on extraction of a surplus value from workers or the human labour capacity. Marx argues that human labour is the sole source of value creation. The production process involves, according to Marx's analysis of the manufacturing process in the nineteenth century, three interacting elements: a) raw materials, b) machines, tools and auxiliary materials, c) human labour power[4]. Although these three elements appear totally symmetric to the capitalist who purchases them, they have different consequences in terms of the value production process. The value of the first two elements is destroyed or depreciated in the production process. However, the value embodied in them is recreated by transferring it to the commodity produced. Thus their total value is preserved and remains constant. On the other hand the last element of the manufacturing process, the human labour, is not destroyed or depreciated as a result of the value production process. Its input to the manufacturing process yields to the creation of new value. Thus, human labour creates surplus value in the overall production process. These two types of values are, therefore, referred to as constant and variable capital, respectively. Their particular mix in a given production process is referred to as the 'organic composition' of capital, and it denotes the ratio between constant and variable capital[5].

It is clear from this analysis that one way of increasing the surplus extracted from the value creation process is to reduce the cost associated with labour involvement in the process. This is indeed the general strategy adopted by capitalism throughout its development, and with the advances in ICT and related factory automation technologies in the second half of the twentieth century, the impact of this strategy has become particularly conspicuous, as will be discussed in the next section.

Marx, however, prophesied that the increase in the organic composition of capital cannot be maintained indefinitely, that is, the value production process cannot be totally free of the human element, as human labour power is the sole source of surplus value. In other words, total automation of the production process would not yield to any surplus value, thus, profits according to Marx. The reason for this, in general terms, is that while productivity increases with the increase in the organic composition of capital, given that the labour power has been made redundant, the purchasing power needed to buy them, thus, average profit per unit of output decreases[5,i]. According to this analysis, therefore, total automation of the production process is not feasible in the framework of the classical mode of the capitalist production system.

## Into the perpetual innovation economy: from surplus value to surplus knowledge

The industrialist who has to compete with the others by cutting costs (primarily labour costs) seems to be trapped in a situation where increased investments in dead capital (machines, technology, etc.) yield less and less returns. It is argued that the industrialist escapes the diminishing returns of technological investment by shifting the focus of the economic activity from the production of goods to the production of innovation, i.e., production of new knowledge for the making of goods[6]. According to this argument, businesses which perpetually modify and develop the knowledge applied to the manufacturing of goods may sustain an increased rate of profits even though the average rate of profit is decreasing by appropriating surplus value from less efficient businesses (who lose their market shares to more efficient ones), and by creating new branches of production, i.e., new products[5,ii]. This new mode of production that capitalism enters is termed as the 'perpetual innovation economy' by Morris-Suzuki[2,6].

The introduction of new products in the market is usually accompanied by an increased demand in human labour employed in information or knowledge intensive sectors of the economy, such as planning, co-ordination, management, administration, design and research and development. The demographic shifts in employment patterns in the last 25 years seem to suggest that while jobs are being lost in other areas of the economy, in knowledge-based sectors there is an increase in the number of people employed[1]. In other words, while automation eliminates the human element in the production plant, it increases in all other departments that precede it[6], supporting Marx's claim that there could be no human-free production in a capitalist mode of production.

Thus, as long as extra employment caused by the creation of new products and new methods of production surpasses the unemployment caused by automation, it is argued that high levels of replacement of human labour by information technologies can be sustained in the framework of the capitalist mode of production over a fairly extended period of time[5,6]. However, it is worthwhile noting here that it is far from certain whether the increase in the number of knowledge based jobs can continue indefinitely in this way. Martin[1], for instance, observes that, the growth rate of information based jobs is slowing down in the recent years and the information sector may be reaching the limit of its ability to absorb displaced workers from other sectors[iii].

## Social production of surplus knowledge

Morris-Suzuki[2] identifies three principal means by which businesses shift the focus of their economic activities towards the production of new knowledge.

- Companies relocate their workers from the manufacturing process to such areas of the production process as planning, co-ordination, management, administration, and research and development. These processes contribute to the production of new knowledge applied in the manufacturing of goods. In this case the company does not directly sell information as a commodity, but uses information to add value to the products manufactured.
- Increasing number of firms begin to specialise in the production and sale of 'producer information', i.e., software, databases, and designs.
- Increasing number of companies are engaged in the production and sale of 'con-

sumer information', i.e., books, periodicals, television programmes, videos, home-computer software, etc.

Well-known peculiar characteristics of the information products should be recalled at this point (summarised in Morris-Suzuki[2]).

- Knowledge once produced can be reproduced and transmitted at very low cost.
- Knowledge is never consumed (depleted).
- Principally, because of the above characteristics of knowledge, it is extremely difficult to maintain monopolies of knowledge indefinitely and privately owned knowledge tends to flow back into the public domain.
- For the above reasons, knowledge can only acquire a price (exchange value) when it is protected by some form of monopoly (e.g. copyright, patents etc.).

The above mentioned characteristics of knowledge/information mean that unless the cycle of production of new knowledge applied to innovative products and production techniques is continually repeated, knowledge used in manufacturing advanced products will surely fall into the hands of the competitors and the profits will diminish rapidly. It should however be noted that much of the knowledge used in creating new products and means of production comes as free social knowledge and appropriated and turned into a source of private profit[2]. This phenomenon was already known in the days of Marx: 'Science, generally speaking, costs the capitalist nothing, a fact that by no means prevents him from exploiting it' (quoted in Morris-Suzuki[2]).

Free social knowledge comes basically in two forms[2]:

- education (including public information and library services, public broadcasting facilities etc.);
- informal social knowledge.

Informal social knowledge includes knowledge of language, culture, traditions etc. which is passed from one person to another through social intercourse and it is the basis of the first type of knowledge mentioned above (i.e., education). Both types of knowledge do not directly cost anything to the entrepreneur. In particular, it is virtually impossible to attach any measurable cost to informal social knowledge, which is passed from parents to children, and through socia interaction in general.

## Social information filtering systems and collaborative information retrieval

Information retrieval is usually thought to be a process of transferring knowledge stored in databases to those who may find it useful. Thus, traditional information retrieval work is mainly concerned with mechanisms that facilitate access to knowledge. However, it has increasingly become obvious that knowledge contained in large databases could also be used as an important resource for production of new knowledge. For instance, Swanson[7,8] demonstrates that by discovering hitherto unnoticed logical connections between a pair of documents, it is possible to create new knowledge. One of the Swanson's examples concerns the case of Raynaud's syndrome and fish oils[7]. These two subjects are treated separately in the medical literature. However, Swanson shows that it is possible to generate a new hypothesis about the possibility of using fish oils in ameliorating the condition by bringing together the information contained in separate documents on these two subjects.

From a more general perspective, Karamuftuoglu[9,10] identifies information transfer and production of new knowledge as two distinct but equally important functions of information retrieval systems. This function of IR systems has, until now, gone largely unnoticed in the information science community with the exception of a few researchers[7,8,11,12,13]. These researchers, however, look at knowledge creation by retrieval, mostly, in terms of discovering unnoticed connections between documents, thus, implicitly conceptualising the knowledge production process mainly as a private labour[iv]. However, it can be argued that knowledge production is fundamentally a collective process and advances in network centric computing make the collective aspect of the knowledge production process highly visible[10].

Network based information systems and services (e.g. the Internet) make collaboration between geographically and/or temporally distant participants increasingly feasible. This aspect of network based information systems could be utilised in the context of knowledge production in IR to facilitate collaboration between distributed participants. One of the main features of the emerging networked information systems is the importance of informal ways of collectively discovering, collecting, distributing and sharing information. Informal social knowledge of the human agents is being increasingly used by such information systems to create new information and knowledge. It is the argument of this paper that such informal and collaborative means of knowledge generation becomes particularly relevant in the context of the perpetual innovation economy discussed above.

The following is an overview of some of the collaborative applications that are being developed on information networks.

### Discussion lists and Usenet newsgroups

Discussion lists enable people with a common interest to exchange views, ideas, and messages and share their knowledge and expertise to solve problems using email. Most email software enables one message to be sent to multiple recipients. There is also dedicated mailing list manager software running on servers. Using a simple email program or a dedicated list manager, it is possible to build a conferencing system that focuses on a single or number of closely related subjects. There are thousands of such discussion lists covering a vast variety of subjects.

Usenet is a distributed conferencing system based on the UUCP and NNTP protocols (Unix to Unix Copy Program, and Network News Transport Protocol, respectively). A message sent to a Usenet group is propagated via a 'server-to-server' distribution scheme to a number of other servers[14]. Anyone with appropriate client software can then read the message, send a reply or post the URL of relevant information elsewhere, by connecting to a local server that carries the list. There are several thousand such lists on the Usenet.

According to Shank[15], interaction on discussion lists and Usenet newsgroups is like discussion in the oral form. However the initial sender of the message, or the starter of the 'thread', has no control over its later development. Anyone interested can 'jump in', yet still all voices can be heard. Shank[15] calls this sort of interaction 'multiloguing'[v]. Shank argues that, people on these discussion groups use a kind of logic known as 'abduction'[16]. This type of logic is characterised by pulling together disparate pieces of information to create new hypotheses. While many of the postings on discussion lists and newsgroups are strictly informational, or ask explicitly for information, many others are directed towards implications of ideas, making connections between disparate pieces of information to propose new meaning claims. This makes discussion on the lists interdisciplinary and inclusive. Since participation in these lists is interdisciplinary not only information but also worldviews are shared, making this a perfect medium for the creation of new knowledge. In Shank's words '... toss out a new insight, and more than likely a new thread will develop'[15].

### Social or collaborative information filtering systems

These are also known as 'recommender systems'. These systems are based on the idea that, in everyday life people rely on recommendations made by other people in deciding whether or not to read a book, see a movie, buy a record and so on.

Social information filtering or recommendation systems simulate and augment word of mouth type of recommendations made in everyday life. These systems rely on the voluntary efforts of several people that relate to each other through networked information systems to discover and evaluate new information sources[17,18,19]. In a typical social filtering system several users provide evaluations of the information objects, the system then aggregates the user judgements and directs to appropriate recipients. The contents of evaluations in such systems can be anything from unstructured textual annotations to binary 'recommended or not' type decisions. Recommendations may be entered explicitly by the user or gathered implicitly by the system[18].

The concept of collaborative information filtering originated with the Information Tapestry project at Xerox PARC[20]. The Tapestry system allows its on-site users to annotate electronic mail messages using free text and/or make binary accept/reject evaluations. Other users of the system can then use, in addition to the contents of the documents themselves, free text annotations rich in information and/or binary accept/reject recommendations to retrieve messages.

Other systems make use of explicit user ratings of the documents to determine groups of users who share similar interests. Based on this information it is then possible to make personalised recommendations. GroupLens[21] filters Usenet newsgroup messages based on past evaluations of the articles. Ringo[22] makes recommendations for music albums and artists based on the same principle. Hill *et al.*[23] reports a system that makes use of past user rankings to select movies. Answer Garden [24,25] is an example of a tool for the creation of information repositories or 'group memories' in organisational settings. It supports the creation of archives, similar to 'Frequently Asked Questions' (FAQ) files, that answer common questions. If the user cannot find an answer in the archives, she/he can route the question to an appropriate human expert through the system.

Maltz and Ehrlich[26] report another system, which makes use of the social approach to information filtering. The system adopts an 'active' approach to information filtering, whereby people send interesting new documents directly to the other users of the system. The users create 'pointers', which are hypertext links to the source documents supplemented with free text comments and contextual information. This information is either emailed to one or more people, or saved in private or public databases. The system takes advantage of a group of people who know each other's interests and are willing to share the discovered information. The benefit of the system increases with the divergence of the documents that are found[26].

This is an important characteristic of social information filtering systems. In a 'content-based' filtering system, documents similar to the ones known to the user of the system are selected, thus new areas that might be of interest to the user can be missed completely. There are no inherent methods for generating serendipitous finds in a content-based filtering system. The system recommends more of what the user already has seen before and indicated liking[26]. In a social information system neither the senders nor the recipients are constrained to pre-determined topics. This is particularly useful in a knowledge production (e.g. research) setting where participants are actively exploring and learning new ideas.

## Research and electronic community systems

Many types of 'knowledge workers', including managers, researchers, and academics, spend a significant part of their time working in groups. To facilitate effective collaboration between group members different types of 'groupware' have been devised. Among these 'Integrated Collaborative Research Systems' (ICRS) is the most relevant for the purposes of this paper. ICRS support various stages of knowledge production activities, such as research. Two examples of such systems are briefly described below.

The Worm Community System (WCS) is an electronic community system for molecular biologists who study the nematode worm Caenorhabditis elegans[27,28,29]. At the height of its development, researchers from about 50 laboratories around the world had used the system. The WCS links various C. elegans databases containing both formal and informal literature and data. It allows the users to retrieve related information held in separate databases concurrently. In addition to retrieving information from different sources, scientists using the Worm Community System can also add their own information to the systems' various databases. New information entered to the system can be made available to a selected portion or the whole of the user community. An important feature of the WCS is that it enables its users to create new links between information stored in separate databases, to create new knowledge. This function of the system allows a user to record links between information objects hitherto held independent of each other. Once a new link is recorded between some information objects, it is available to the other users of the system when one of the linked objects is retrieved at a later time. In this way, the WCS brings together knowledge production and transfer functions of information retrieval noted earlier.

The Arizona Analyst Information System is an information management system for 'research collaboration'[30]. It is designed to support collaborative research on IT transfer and international technological trend and policy analysis. The system integrates many of the functions needed to support activities in a typical research cycle, including entry, indexing, retrieval, analysis and writing of large volumes of textual information. The user group comprises some 20 researchers with different levels of expertise and different types of information needs. To enable effective collaboration between group members the system aims to promote communication and sharing of information, capture and distribution of group expertise, capture and maintenance of an organisational or group memory. One of the tools in the Arizona system that serves some of the above objectives is a 'HOT' file. A 'HOT' file is a special kind of file that acts as a group information repository. Any group member can post a document or a piece of information to one of the 'HOT' files when the researcher believes that it could be of interest to another member of the group. 'HOT' files have characterisations associated with them, which enables the automatic routing of a document to the relevant user or users of the system, thereby creating a portion of the group's collective memory that is necessary for effective communication between the group members.

From the above discussion it is clear that collaborative applications emerging on networked systems aim to harness the expertise and knowledge embodied in human agents to produce new information and knowledge. The success of many of these systems depends on the skills acquired by the individuals through the process of socialisation (e.g. language, ability to associate with others etc.), and the bonds that emerge between the members of social groups. Social bonds develop mutual dependence and necessitate collaboration and sharing between the group members.

At this point, it is appropriate to consider the role of the collaborative technologies emerging on information networks in the context of the perpetual innovation economy discussed in the preceding pages. Perpetual innovation economy requires creative individuals who work in teams to generate innovative designs and products rapidly. To achieve this objective, the old model of the workplace or the factory as a facility to reproduce blueprints has been abandoned. In information based economies the factory is re-conceptualised and managed as a 'learning environment' which continuously harnesses the human capability to be creative. The fundamental unit in a creative learning and innovation environment is the team or social group. The most successful production facilities around the world work on this principle[31]. For instance, it has been noted that, Japanese industry encourages the social and collective aspect of work to harness a team's ability to innovate and generate new solutions[31]. Such management techniques explicitly seek to re-socialise the workplace and manage the social aspect of the innovation process. This is in direct contrast to the older management techniques based on Taylorism/Fordism, which divides the production process to individual efforts. Individual workers in the Taylorist model are isolated and compartmentalised[31]. Not unlike users and information objects (documents, keywords, descriptor terms, etc.) in most traditional retrieval systems and mechanisms.

Libraries and electronic databases are traditionally seen as depositories of objective knowledge. Consequently, their role is seen as an effective transfer of knowledge stored in databases to the users[32]. In most traditional retrieval systems, documents and other information objects are divided into atomised or isolated units[33]. Users are similarly treated as private individuals, isolated and compartmentalised[10]. This model of information storage and retrieval is true even for more recent keyword-based search engines that are found on the World Wide Web. The current model of the search engines on the web excludes the users who are concurrently logged on the same search service (this is true more generally for the model of the WWW itself). While documents and other information objects on the web are linked the users are treated as isolated agents, notwithstanding the fact that the users of the web (and the Internet generally) are in reality linked by means of networked servers.

We have seen, from the above discussion of the perpetual innovation economy, that social means of gathering, sharing, and using information is becoming increasingly important as a way of production of new knowledge applied to manufacturing of goods. It is therefore possible to argue that for the information profession to hold an important stake in the information economy, it is necessary to re-conceptualise the role of information practitioners as producers of information/knowledge as well as neutral guardians of knowledge, and the retrieval tools and systems they devise as tools for production of new knowledge, as well as tools for effective storage and retrieval of knowledge. This requires re-conceptualisation of information seeking, retrieval and production as a collective and collaborative process, and not individualistic[10,34]. To this end, it should be useful to study and learn from collaborative systems that are being developed on information networks and the Internet, some of which are reviewed in the preceding pages.

## Conclusion

The overall objective of this article is to show the importance of social knowledge and social (collaborative) means of gathering, distribution and use of information for information/innovation based economies. In a perpetual innovation economy, knowledge or information embedded in commodities, not the physical material from which commodities are made, becomes the main source of profit. However, the peculiar character of information/knowledge means that privately owned knowledge tends to flow back into the public domain. Conversely, knowledge underlying production processes and produced goods is drawn, to a large extent, from social stock of knowledge. The peculiarity of information/knowledge necessitates continuous renewal and modification of products and production techniques. Innovative and, more generally, creative acts are fundamentally collective and collaborative processes. In a perpetual innovation economy, it is therefore necessary to cultivate, manage and harness the social skills and social knowledge of the human work force to maintain constant product renewal and development. It is important to reflect on the role of information professionals and information management tools and techniques in this context.

Traditionally, the role of information professionals (such as librarians) has been seen to preserve knowledge contained in documents and devise effective mechanisms of access to knowledge. From this perspective, users and information objects are seen as isolated and compartmentalised entities. It is obvious that information practitioners (and information storage and retrieval tools devised by them), perceived merely as guardians and facilitators of knowledge, do not have much to aspire to in the context of the perpetual innovation economy. At this point, knowledge production potential of information storage and retrieval systems should be appreciated. This aspect of retrieval systems is only barely acknowledged in the information science community. However this is an important aspect of such systems. The knowledge production function of retrieval systems becomes particularly significant when it is considered in relation to innovation based economic systems.

Creative acts are not individualistic but collective processes. In order to facilitate creativity, it is necessary to cultivate social relations and harness knowledge embedded in such relations. Emerging collaborative systems on computer networks, such as the Internet, make it possible to devise work environments that are conducive to the development and cultivation of social practices. Some of the collaborative applications being developed on computer networks are reviewed in earlier sections of this article. Information retrieval systems designers and practitioners may find it useful to study such systems in order to conceive retrieval mechanisms that facilitate the production of new knowledge as well as transfer of the existing knowledge.

It is hoped that by putting information retrieval in the context of the latest development in economic relations, in particular the perpetual innovation economy, the knowledge production potential of information storage and retrieval systems and tools becomes more widely acknowledged and accepted among information practitioners. Although it is not one of the objectives of this article, it may be, nevertheless, useful to note here briefly that information professionals and designers of retrieval systems should also reflect on the consequences of a close alliance between the information profession and the requirements of the innovation economy. Developing, implementing, and using systems that cultivate and manage social relations and knowledge solely according the requirements of the economic order may result in too narrow a view of the role of knowledge and social relations in industrial production and more generally in the society as a whole.

## References

1. Martin, S. 'Information technology, employment, and the information sector: trends in information employment 1970-1995.' *Journal of the American Society for Information Science* 49(12), 1998, pp.1053-1069.

2. Morris-Suzuki, T. 'Capitalism in the computer age and afterword.' In Davis, J., Hirschl, T. and Stack, M. (eds.) *Cutting Edge: technology, information, capitalism and social revolution.* (London: Verso, 1997, pp.57-71.)

3. Davis, J., Hirschl, T. and Stack, M. *Cutting Edge: technology, information, capitalism and social revolution.* (London: Verso, 1997.)

4. Caffentzis, G. 'Why machines cannot create value; or, Marx's theory of machines.' In Davis, J., Hirschl, T. and Stack, M. (eds.) *Cutting Edge: technology, information, capitalism and social revolution.* (London: Verso, 1997, pp.29-56.)

5. Carchedi, G. 'High-tech hype: promises and realities of technology in the twenty-first century.' In Davis, J., Hirschl, T. and Stack, M. (eds.) *Cutting Edge: technology, information, capitalism and social revolution.* (London: Verso, 1997, pp.73-86.)

6. Morris-Suzuki, T. 'Robots and capitalism.' In Davis, J., Hirschl, T. and Stack, M. (eds.) *Cutting Edge: technology, information, capitalism and social revolution.* (London: Verso, 1997, pp.57-7113-127.)

7. Swanson, D.R. 'Two medical literatures that are logically but not bibliographically connected.' *Journal of the American Society for Information Science* 38, 1987, pp.228-233.

8. Swanson, D.R. 'Online search for logically-related noninteractive medical literatures: a systematic trial-and-error strategy.' *Journal of the American Society for Information Science* 40, 1987, pp.356-358.

9. Karamuftuoglu, M. 'Semiotics of documentary information retrieval systems.' In Ingwersen, P. and Pors, N.O. (eds.), *Information Science: Integration in Perspective. Proceedings of CoLIS 2, Second International Conference on Conceptions of Library and Information Science.* (Copenhagen: Royal School of Librarianship, 1996, pp.s85-97.)

10. Karamuftuoglu, M. 'Colaborative information retrieval: toward a social informatics view of IR interaction.' *Journal of the American Society for Information Science* 49(12), 1998, pp.1070-1080.

11. Bawden, D. 'Information systems and the simulation of creativity.' *Journal of Information Science* 12(5), 1986, pp.203-216.

12. Davies, R. 'The creation of new knowledge by information retrieval and classification.' *Journal of Documentation* 45(4), 1989, pp.273-301.

13. Gordon, M.D. and Lindsay, R.K. 'Toward discovery support systems: a replication, re-examination and extension of Swanson's work on literature-based discovery of a connection between Raynauds and fish-oil.' *Journal of the American Society for Information Science* 47, 1996, pp.116-128.

14. December, J. 'Units of analysis for Internet communication.' *Journal of Communication* 46, 1996, pp.14-38.

15. Shank, G. 'Abductive multiloguing, the semiotic dynamics of navigating the net.' *The Arachnet Electronic Journal on Virtual Culture* 1, 1993.

16. Buchler, J. (ed.) *Philosophical writings of Peirce*. (New York: Dover, 1955.)

17. Hill, W., Stead, L., Rosenstein, M. and Furnas, G. 'Recommending and evaluating choices in a virtual community of use.' In *Human Factors in Computing Systems (CHI '95) Conference -Proceedings* 1 (New York: ACM, 1995, pp.194-201.)

18. Resnick, P. and Varian, H.R. 'Recommender systems.' *Communications of the ACM* 40(3), 1997, pp.56-58.

19. Starr, B., Ackerman, M.S. and Pazzani, M. 'Do-I-Care: a collaborative web agent.' In *Human Fac-*

*tors in Computing Systems (CHI '96) Conference – Proceedings.* (New York: ACM, 1996, pp.273-274.)

20. Goldberg, D., Nichols, D., Oki, B.M. and Terry, D. 'Using collaborative filtering to weave an information tapestry.' *Communications of the ACM* 35, 1992, pp.61-70.

21. Reisnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J. 'GroupLens: an open architecture for collaborative filtering of netnews.' In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work*, 1994, pp.175-186.

22. Shardanad, U. and Maes, P. 'Social information filtering.' In *Human Factors in Computing Systems (CHI '95) Conference – Proceedings* 1 (New York: ACM, 1995, pp.210-217.)

23. Hill, W., Stead, L., Rosenstein, M. and Furnas, G. 'Recommending and evaluating choices in a virtual community of use.' In *Human Factors in Computing Systems (CHI '95) Conference – Proceedings* 1 (New York: ACM, 1995, pp.194-201.)

24. Ackerman. M.S. and McDonald, D.W. 'Answer Garden 2: merging organizational memory with collaborative help.' In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work,* 1996, pp.97-105.

25. Ackerman. M.S. and Malone, T.W. 'Answer Garden a tool for growing organizational memory.' *SIGOIS Bulletin* 11, 1990, pp.31-39.

26. Maltz, D. and Ehrlich, K. 'Pointing the way: active collaborative filtering.' In *Human Factors in Computing Systems (CHI '95) Conference – Proceedings* 1 (New York: ACM, 1995, pp.202-209.)

27. Chen, H., Yim, T. and Fye, D. 'Automatic thesaurus generation for an electronic community system.' *Journal of the American Society for Information Science* 46, 1995, pp.175-193.

28. Poole, R. 'Beyond database and e-mail.' *Science* 262, 1993, pp.841-843.

29. Schatz, B. 'Building an electronic community system.' *Journal of Management Information Systems* 8, 1991/1992, pp. 87-107.

30. Chen, H. 'Information management in research collaboration.' *International Journal of Man Machine Studies* 36, 1992, pp.419-43.

31. Kenney, M. 'Value creation in the late twentieth century: the rise of the knowledge worker'. In Davis, J., Hirschl, T. and Stack M. (eds.) *Cutting Edge: technology, information, capitalism and social revolution* (London: Verso, 1997, pp.87-102.)

32. Radford, G.P. 'Positivism, Foucault, and the fantasia of the library: conceptions of knowledge and the modern library experience.' *Library Quarterly* 62(4), 1992, pp.408-424.

33. Karamuftuoglu, M. 'A deleuzoguattarian framework for understanding information systems: the case of document retrieval systems.' In Gupta, J.N.D. (ed.) *Proceedings of the Americas Conference on Information Systems* (Association for Information Systems, 1997, pp.882-884.)

34. Hjørland, B. *Information seeking and subject representation: an activity-theoretical approach to information science.* (London: Greenwood Press, 1997.)

## Notes

[i] This is a very schematic overall description of the relations between human labour, automation, and the capitalist mode of production and the ways in which they interact with each other. The reader is referred to Carchedi[5] and Morris-Suzuki[2,6] for a more detailed analysis.

[ii] The summary of the appropriation of surplus value by the introduction of technological innovation and competition is necessarily schematic here. The interested reader should refer to Carchedi[5], Morris Suzuki[2,6].

[iii] Martin[1] also notes that growth rate is higher in non-routine information handling jobs which require substantial education and training than in routine/clerical information handling jobs which require lower skills.

[iv] As this sort of 'discovery' does not require collective effort in a fundamental way (see Karamuftuoglu[10]).

[v] After, dialoguing, monologuing etc.

# Internet, intranet, extranet: managing the information bazaar

*Ian Watson*
*Information Services Manager, Scottish Media Newspapers, 195 Albion Street, Glasgow G1 1QP*
*Email: iwatson@cims.co.uk*

**Abstract** *The web has brought to the business world an inexpensive way of distributing corporate information to employees, business partners and customers. The term intranet describes a closed computer network built upon World Wide Web technology in which access is restricted to a particular group of users, typically employees of a company. The term extranet is used when access is extended to a privileged user group: customers and suppliers for example.*

*This paper examines the role of the World Wide Web in the Research Library at Scottish Media Newspapers, in particular how it complements the portfolio of traditional online sources, CD Roms and books. It also describes the extent to which searching has moved from the intermediary to the end user.*

*In the mid 1990s, just as the web was beginning its spectacular rise to fame, Scottish Media Newspapers developed a browser-based interface to the in-house database of newspaper stories, a process that led to the creation of a corporate intranet. The creation of the intranet has been accompanied by a redefining of staff roles in an attempt to change the image from passive librarian to dynamic research specialist. The paper concludes by examining the role of the information professional/librarian in a world where end users will have a vast range of information from internal and external sources at their fingertips.*

## Introduction

Although the worldwide computer network known as the Internet has been around since the late 1960s, until the early 1990s it was a world largely inhabited by the relatively small group of scientists who had access to expensive mainframe computers. During the 1970s and 1980s online information hosts such as Data Star and Dialog were creating a quite separate electronic information world, one inhabited mostly by librarians and information specialists. The arrival of the World Wide Web (WWW or 'the web') forced these two worlds into collision, in the process thrusting online information into the mainstream of popular culture and probably changing for ever the nature of the information business.

The WWW is simply a friendly graphical user interface (GUI) to the technologies and protocols that underpin communication between computers on the Internet. As the number of web sites continues to explode so does the number of search services seeking to bring some order to this great electronic information bazaar. Services such as HotBot, Yahoo!, Excite, and Alta Vista are useful but experienced online searchers may find them rather clumsy in comparison to the highly structured services provided by traditional online hosts. Simple comparisons with traditional online hosts, however, fail to recognise that web technology is revolutionising the way information systems work. The tools that have emerged for searching the web, in particular the ability to build hypertext links to related documents, can be highly effective in building corporate information networks, or intranets as they are commonly known.

It is difficult to understate the impact of the web. Hardly known five years ago, the characteristic *http://www.xxx.com* is now everywhere. Some users have very high expectations of the web, believing that everything imaginable can be found there while the cynics see it as an amusing toy or an enormous time-wasting creation. Using Scottish Media Newspapers as a case study, this paper looks beyond this hype to find out what the web means for those of us working with electronic information.

## The web and the Internet

Part of the Scottish Media Group, Scottish Media Newspapers (SMN) publishes *The Herald* and the *Evening Times*. Established in 1783, *The Herald* (formerly *Glasgow Herald*) is one of the oldest English language newspapers in the world and is Scotland's leading quality broadsheet. The *Evening Times* is Scotland's best-selling evening newspaper, circulating mostly in the Glasgow area. The web has added a new tool to the research library's armoury of traditional online hosts, CD Roms and reference books and is now heavily used both by end-users and intermediaries, particularly to find ephemeral information.The web is most useful for primary sources of information such as Parliamentary debates, timetables for Parliamentary business, statistical material, political party manifestos etc. It is also very useful for anecdotal information about pop groups, the film industry, footballers and football teams. The various services such as AltaVista, Excite, Lycos and Infoseek have their own particular approaches to managing the web. HotBot is good for exhaustive trawling, but when searches return thousands of sites the effectiveness of the service becomes questionable. Recent arrival Google uses the number of links to a site to rate its importance. Ask Jeeves, on the other hand, is an example of a tool that tries to imitate the experienced librarian. Ask a question and it will respond with several more in an attempt to elicit more accurate details of what you really want. For example 'Who is the Prime Minister of Britain?' will generate the following questions

1. 'Do you want British heads of State?'
2. 'Where can I hear English MIDIS?'
3. 'Do want an episode guide to Yes Prime Minister?'

Choose the first and it responds with a list of all British Prime Ministers this century and, thrown in for good measure, lists of party leaders, Foreign Secretaries etc. It performs quite well on highly factual questions but is less certain when there is ambiguity of meaning in the search terms.

The interesting feature about Ask Jeeves is that it uses natural language processing to understand and match users' questions to an extensive knowledge base. The knowledge base consists of thousands of question templates and millions of researched answer links to web sites compiled by research staff. The publicity material proudly claims supremacy over other search services precisely because 'the knowledge base is built by humans, not by software spiders'. Far from putting information intermediaries out of business, Ask Jeeves appears to offer a new role in building the knowledge base for particular domains. Search services, however, are essentially aimed at the consumer market, and it is interesting to examine the role of the traditional online hosts, faced with a more information-aware world, clamouring for cheaper, if not free, products.

## The web and online hosts

As the web browser has become an ubiquitous feature in offices and homes, several traditional online hosts now provide web front ends, a trend which is to be encouraged. At Scottish Media Newspapers we are keen to provide end-user access to online services but maintaining a plethora of proprietary Macintosh or Window-based interfaces carries a substantial support overhead. For example we subscribe to Reuters Business Briefing which is installed on a number of Macintoshes and Windows machines. Installing proprietary software, providing end user training and troubleshooting all require time. This overhead is reduced when all users use the same browser to access all online sources. The main overhead then becomes training in the use of the service rather than means to get there. In short, the web simplifies access.

This simplified access will enable end users to undertake routine searches on familiar services. For example Bureau Van Dijk sells access to the UK electoral register via the web while Advance Media Information offers a very useful service for getting information about forthcoming events, and both these services are very simple to use. Not everyone on every occasion, however, will feel that it is appropriate to do their own searching. When time is short, the search is complex, or the source unfamiliar, they might turn to the expert searcher. To take an example from the world of do-it-yourself: home decoration products have never been easier to use, but there are many situations in which paying an expert is the easiest and wisest option!

### Towards an intranet

The term 'intranet' refers to a computer network built upon Internet technologies in which access is restricted to a particular group of users, typically employees of a company. When that network is expanded to include other authorised users the term 'extranet' is sometimes used. In this way the web has brought to the business world an inexpensive way of distributing corporate information to employees, business partners and customers. At its simplest an intranet offers a client-server cross-platform solution to information management problems. In other words it provides a uniform way to access information regardless of where or on what platform it resides and regardless of what kind of computer is sitting on the user's desktop. In its most basic form an intranet is made up of just three pieces of software: a web server, web browser, and firewall to keep unauthorised visitors out. A fourth component – a search engine – is also essential as the volume of information to be managed grows.

For many years the primary activity of the typical newspaper library was the maintenance of newspaper cuttings files. In the 1980s the increased availability and sophistication of full text retrieval software paved the way for the creation of digital text archives. In 1989 a digital archive of the editorial content of *The Herald* and *Evening Times* came online using Basis retrieval software from Information Dimensions. In theory this should have provided a better service to journalists by removing the reliance on cumbersome cuttings files which were always subject to the vagaries of classification procedures, mis-filing and even loss. The archives would now be searchable from the journalists' desks which ought to have been a very welcome development. In practice, however, the command-line interface was never popular, except with a very small number of journalists, and attempts to simplify the interface with menus and prompts were largely unsuccessful in overcoming the inherent difficulties of using arcane search commands. This mirrored the problem faced by online database hosts in their attempts to make their systems attractive to the end user rather than the professional intermediary. The problem was that the user interface – quite acceptable to the information professional – was visually dull and not very intuitive for the occasional user.

By 1995 our newspaper production system had migrated to a Quark-based system and our journalists were working on Macintoshes. Around the same time Information Dimensions released a new version of Basis which incorporated a web server. With the web server sitting on top of the Basis database, all that was needed was a browser on the user's desktop machine. Many journalists were already familiar with Netscape from their use of the newly emerging World Wide Web, so they were not confronted with an unfamiliar and non-intuitive command-line interface. Later it was very easy to extend access to Scottish Television – another company with the Scottish Media Group – where a Windows environment prevails and Internet Explorer is the preferred browser. Diagram 1 shows the home page.

Desktop access to the archive using a friendly GUI has not only been a major boon to the journalist, it has also freed library staff from a certain amount of routine searching, allowing them to concentrate on the management of information resources. With the technology in place it was a relatively simple step to create a dynamic information system. Customising the browser required only a minimal knowledge of HTML and the first step was to add information and advice files:

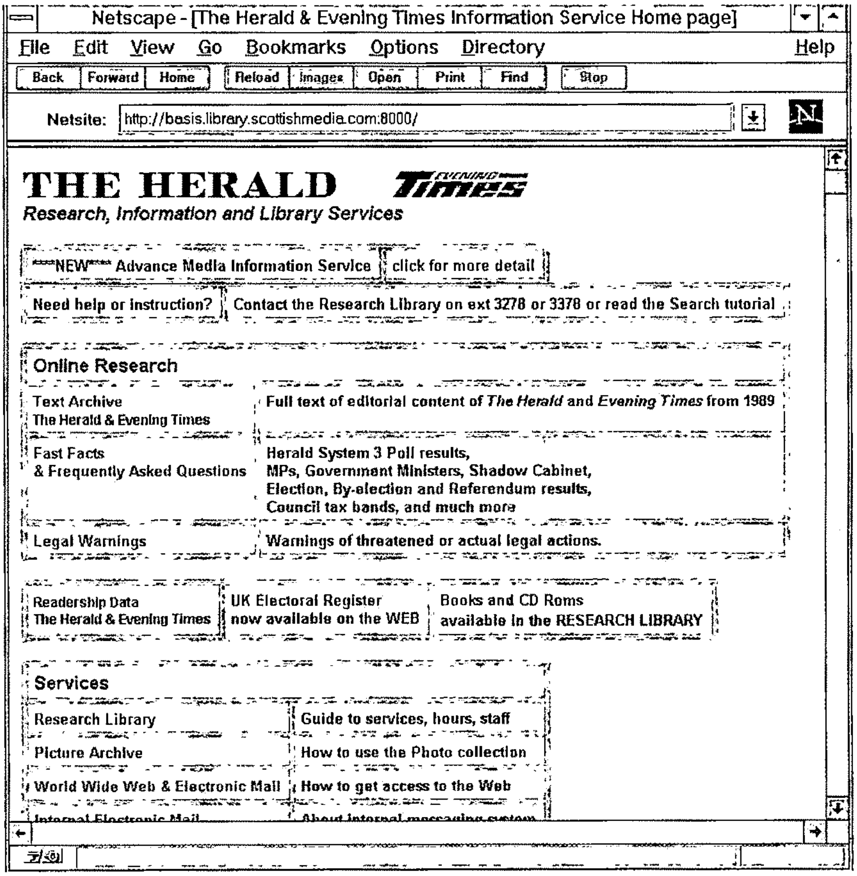
- guide to information services;
- catalogue of books and CD ROMs;
- newspaper sales and readership information;
- announcements.

Information held centrally can be updated once without worrying about replacing out-of-date paper copy. Research library staff can concentrate on organising information confident that there is a very simple method of dissemination. For example, a 'fast facts' database has been established to store information traditionally kept by library staff in notebooks or card files. Frequently asked for information such as the date the first man landed on the moon or the number of people killed in various disasters is now more readily available. Tables of election and devolution referendum results, lists of MPs and Government Ministers, once pinned to the wall for rapid access, are now on the intranet. As the librarian emerges into a more proactive role, the intranet offers a highly effective way of pushing information at the user. There is nothing new in compiling such information packs. What is new is the distribution channel.

A very useful feature of the web is that every document has a URL and this is true whether we are looking at the WWW or an intranet. Basis software automatically gives every document in the database a URL which makes it possible to create hypertext links between documents in different databases. For example, from a document in the Fast Facts database recording the date when Glasgow Celtic became the first British football team to win the European Cup we can create a link to a selection of relevant newspaper stories. From stories in the archive we can also create links to a database containing details of legal warnings relevant to these stories.

While intranet technology is relatively simple, its management can cause headaches for information technology departments faced with the administration of a proliferation of devices and addresses. Intranets also have the potential to significantly increase traffic which can cause bandwidth problems. Care has to be taken therefore to make sure that the information available can be handled by the network. Some technology managers

*Diagram 1. Home Page*



http://basis.library.scottishmedia.com:8000/

**THE HERALD** *Evening Times*

*Research, Information and Library Services*

| | |
|---|---|
| ***NEW*** Advance Media Information Service | click for more detail |
| Need help or instruction? | Contact the Research Library on ext 3278 or 3378 or read the Search tutorial |

Online Research

| | |
|---|---|
| Text Archive<br>The Herald & Evening Times | Full text of editorial content of *The Herald* and *Evening Times* from 1989 |
| Fast Facts<br>& Frequently Asked Questions | Herald System 3 Poll results,<br>MPs, Government Ministers, Shadow Cabinet,<br>Election, By-election and Referendum results,<br>Council tax bands, and much more |
| Legal Warnings | Warnings of threatened or actual legal actions. |

| | | |
|---|---|---|
| Readership Data<br>The Herald & Evening Times | UK Electoral Register<br>now available on the WEB | Books and CD Roms<br>available in the RESEARCH LIBRARY |

Services

| | |
|---|---|
| Research Library | Guide to services, hours, staff |
| Picture Archive | How to use the Photo collection |
| World Wide Web & Electronic Mail | How to get access to the Web |
| Internal Electronic Mail | About internal messaging system |

are reported to be concerned that bandwidth for vital business applications is being squandered by less than vital intranet data. The key problem is managing a vital resource called information and there has to be an awareness that an intranet should not become an information free-for-all. Rather it is a tool to help ensure that information is available to those who need it when they need it.

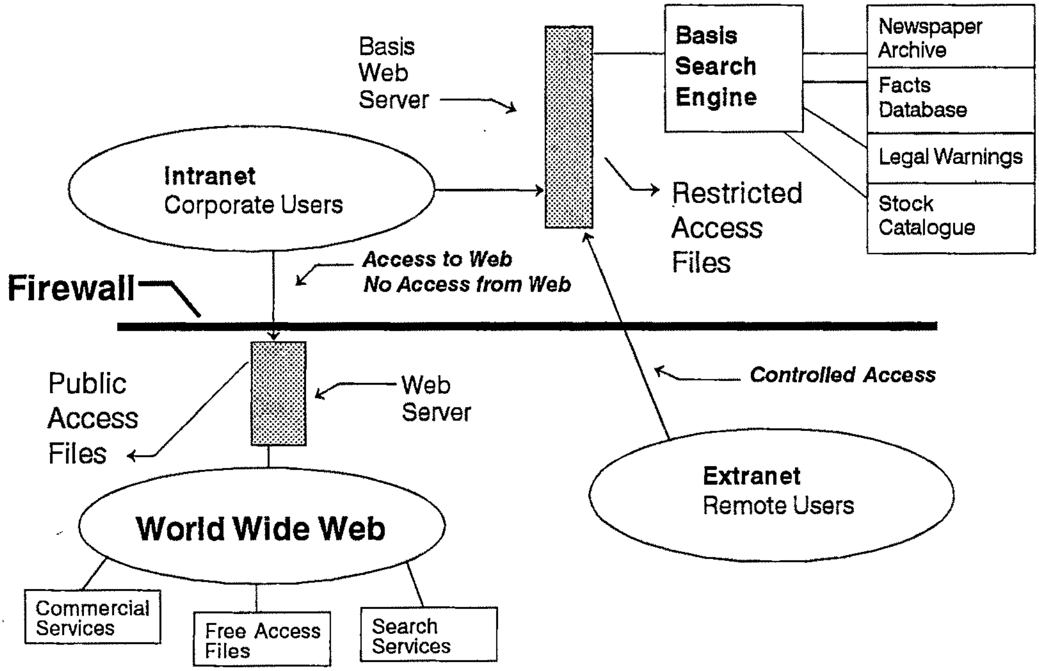
For information professionals the intranet offers a wonderful range of technologies on which they can develop information management skills. At Scottish Media Newspapers we see it as a flexible tool which will enable us to provide a better and more responsive service. It helps cast off the image of the librarian as the custodian of dusty old cuttings and the low status which accompanied such a role. The extranet has allowed us to raise our profile within the Group, which itself profits by deriving more value for money from the investment in information services. Diagram 2 shows our users with access to internal databases and information files and link through a firewall to a server which can access the web. Other users within the Scottish Media Group have controlled access, via ISDN line, to the Basis web server which allows them to search the databases.

So far the intranet has proved very popular with journalists as a simple way of accessing information held in the text archive. Much of the popularity derives from simplicity of use in the same way that use of the Internet has been stimulated by the arrival of simple GUIs. It is noticeable also that the traditional online vendors now see the web as the way of at last achieving the holy grail of end-user searching. Technically there is no reason why the intranet cannot provide a seamless link to an external source, subject to appropriate security safeguards. The librarian will have an important role at the centre of this as an expert in information sources, ensuring that the intranet maintains its currency in a world of expanding and ever-changing information sources.

## The future

The web is not just about finding information: it is about publishing, broadcasting, establishing networks and developing interactive serv-

*Diagram 2. Internet, intranet and extranet*



ices. It has revolutionised the way information systems work and we are likely to see the evolution of custom built web-based services that are hybrids of Internet, intranet and extranet.

We often hear the hype that the WWW has made redundant all the years of work embodied in the creation of structured databases and commercial information services and that it has made information searching skills redundant. Information may well become cheaper than hitherto but the belief that the web, by improving accessibility, will make it free is misguided. Services such as Yahoo! are free to the end user but there is no reason why the information it signposts should be free. Nor should we assume that value-added products such Chemical Abstracts or Medline ought to be free. Search skills are likely to be required when the simple search engines cannot find a specific nugget of information buried in the terrabytes. Yahoo! employs editors and librarians to catalogue and organise the sources it references. Ask Jeeves uses knowledge bases compiled by human beings. The question of provenance is also vital. On the web everyone can be a publisher but how do you know the good from the bad publishers?

There is also the more negative kind of hype which suggests that the web is a passing fad, a plaything and time waster. Information specialists and librarians worked closely with online hosts during the 1970s and 1980s to secure improvements in front ends and search engines. They can play important roles in the future as content managers, using web technologies to customise services to meet the needs of their end-users.

## Web addresses of sites, engines and software mentioned

Information Dimensions *http://www.idi.oclc.org*
Ask Jeeves *http://www.askjeeves.com*
Yahoo! *http://www.yahoo.com*
HotBot *http://www.hotbot.com*
Alta Vista *http://www.altavista.digital.com*
Google *http://www.goggle.com*
Advance Media Information *http://www.amiplan.com*
UK Electoral Register *http://cameo.bvdep.com*

# A national survey of the citizenship information needs of the general public

*Rita Marcella and Graeme Baxter*
*School of Information and Media, The Robert Gordon University, Faculty of Management Building, Garthdee Road, Aberdeen AB10 7QE.*
*Email: r.marcella.rgu.ac.uk*

**Abstract** *This paper reports the results of a survey of information needs and information seeking behaviour of a national sample of the UK population. The survey was the first stage of the Citizenship Information project, funded by the British Library Research and Innovation Centre. In total, 1294 responses were received giving a valid and demographically representative response rate of 45.7%. Major findings include: that the majority of respondents had sought information in the past (59.4%) and that an even greater number predicted a future need for information (78.4%). Over three quarters of respondents said that they would use public libraries and between half and three quarters would approach Citizens Advice Bureaux, Post Offices, Government departments or family and friends. Face-to-face communications and reading a book were the most popular means of accessing information, but a wide variety of other preferred options were cited. Only a small proportion expressed a preference for using a computer to seek information, and there was a clear emphasis on public libraries as an appropriate location for accessing electronic information. A highly significant majority (79.2%) believed that access to information was very important for exercising their rights as a citizen.*

## Introduction and background

This paper reports the results of the first stage of the Citizenship Information research project, funded by the British Library Research and Innovation Centre. The main aims of this project are to: investigate the extent to which members of the UK public have expressed or unexpressed needs for citizenship information; explore their preferred routes to the acquisition of such information; and explore the suitability and approachability of the public library, among other agencies, for the user seeking citizenship information. Stage one of the project consisted of a questionnaire-based survey of a sample of the UK population, designed to gather preliminary data on their use of and need for citizenship information.

In the design and execution of the project, the authors drew upon a range of existing research and published literature. This has included the work of Marshall[1], the National Consumer Council[2] and the Policy Studies Institute[3] on the role of information in citizenship and the distinction between the *citizen consumer* and the *active or participant citizen*. There were also links with some observers' definitions of community information, including those of the Library Association[4] and Donohue[5]. Prior research on information needs, including that of Bruce *et al.*[6], Tinker *et al.*[7] and most importantly Warner *et al.*[8] and the University of Sheffield[9], has also been relevant. Professional and representative bodies, such as the Library Association[10] and the Library and Information Commission[11] have argued the importance of public libraries providing citizenship information; while commentators such as Usherwood[12] and Swash and Marsland[13] have predicted a rise in public expectation of public libraries in terms of citizenship information provision.

The Citizenship Information project has been ongoing at a time of great change in terms of information policy formulation and development in the UK. A non-statutory *Code of Practice on Access to Government Information*[14] was introduced in 1994; while the *Freedom of Information White Paper*[15] was published in 1997 with a view to establishing a general statutory right of access to official records and information. The 1998 *Crown Copyright Green Paper*[16] suggested the possible abolishment of Crown Copyright and the placing of all Government material in the public domain.

The use of Information and Communication Technologies (ICTs) to enable the exchange of citizenship information has also been the focus of much attention. The *government.direct* Green Paper[17], for example, was issued as a prospectus for the electronic delivery of Government services to citizens and the business community; the Central Computer and Telecommunications Agency's online Government Information Service[18] attempts to provide a coordinated single point for information produced by Government departments and agencies; while the UK Citizens Online Democracy (UKCOD)[19] service has conducted some interesting experiments in electronic democracy. Plans are also in place to establish an electronic network linking public libraries – *The People's Network*[20] – which the Government believes will play a central role in delivering its wider objectives for the role of technology in society. The Parliamentary Office of Science and Technology has issued an *Electronic Government* report[21] which discusses the benefits and issues arising from these and other initiatives.

## Methodology

The stage one questionnaire was disseminated across all 13 Government office regions of the UK. Overall, 1294 questionnaires were completed and returned, giving a 45.7% response rate. Each of the regions achieved a response rate between 35.5% and 54.5%, apart from Northern Ireland at 75% and Wales at only 14.7%.

Questionnaires were handed to users in public libraries, Citizens Advice Bureaux and other independent advice agencies. A much better response was achieved in the public libraries (69.4%) than in the CABx (40.5%) and other advice and information agencies (8.3%). With over 75% of the completed questionnaires having been distributed in public libraries, survey results may be somewhat biased towards the opinions of public library users: the Library and Information Commission's report *New Library: The People's Network*[20] indicated that libraries are used by only 58% of the UK public. The survey response does not indicate, however, any groups excluded or poorly represented in the sample.

*Table 1. Shows the gender and age representation of the sample.*

| Age Group | Gender | | | |
|---|---|---|---|---|
| | **Male** | **Female** | **Not specified** | **Totals** |
| Under 15. | 10 | 21 | – | 31 (2.4%) |
| 15-19 | 37 | 61 | 1 | 99 (7.7%) |
| 20-29 | 126 | 120 | 1 | 247 (19.1%) |
| 30-44 | 177 | 190 | 1 | 368 (28.4%) |
| 45-54 | 90 | 111 | – | 201 (15.5%) |
| 55-64 | 82 | 76 | 1 | 159 (12.3%) |
| 65-74 | 71 | 54 | 3 | 128 (9.9%) |
| 75 or over | 33 | 16 | 3 | 52 (4.0%) |
| Age not specified | 4 | 4 | 1 | 9 (0.7%) |
| **Totals** | **630 (48.7%)** | **653 (50.5%)** | **11 (0.9%)** | **1294 (100%)** |

In total, 100 (7.7%) of the respondents belonged to an ethnic minority group. 55.5% of the respondents were economically active (i.e. working or looking for work). With regard to the economically inactive respondents in the survey, 14.1% were students. Respondents to the survey come to a greater extent proportionally from the professional and managerial categories than for the UK as a whole.

There were some interesting differences in status between the respondents in the three types of organisation involved in the survey: the proportion of jobseekers in the CABx (22.5%) and the other advice agencies (21.2%) was effectively twice that in the public libraries (10.9%), which perhaps reflects the concerns this group would have about obtaining information and advice; while the public library student percentage (16.0%) was virtually double those in the CABx (8.7%) and the other agencies (6.1%).

123 respondents (9.5%) indicated they were disabled in some way. The proportion of disabled respondents in the CABx (15.0%) and the other advice agencies (27.2%) was significantly greater than in the public libraries (6.9%).

## Survey results

Respondents were asked to give an example of an occasion on which they had been required to look for information to help them make a decision, solve a problem, or understand something a little better; 769 respondents (59.4%) gave an example. They indicated a wide range of situations in which a need had arisen. Information for educational purposes was cited most frequently, by over a quarter of the respondents, followed by leisure and recreation information (16.9%). Health care information, information on welfare benefits eligibility, and legal information were also prominent. When subsequently presented with a list of 23 subjects, 1100 (85%) had wanted to find out more about at least one of the categories. The top six responses remained: information on education, leisure, health care, welfare benefits, legal issues and employment.

There were some significant demographic differences: for example, 42.1% of the female respondents had indicated a past need for educational information, compared with 31.7% of male respondents: male respondents, meanwhile, had had more interest in technology and communications (+7.0) and political information (+6.2). For certain types of information (e.g. local government, health care, taxation, financial matters, legal issues) the percentages for those aged 19 or under were decidedly lower than the other age groups. Welfare benefits information had been needed in the past by 36.7% of those respondents currently seeking work and by 29.1% of those who were running a home, but only by 17.5% of the retired and 16.5% of the students. Geographic variations were also prevalent if inexplicable. For example, only 9.2% of East Midlands respondents cited a past need for housing information, compared with 31.1% of those in the South East. Ethnic minority respondents demonstrated a greater need for immigration and nationality information and information on equal rights and discrimination. Disabled respondents had more frequently sought information on welfare benefits and health care. Finally, there appeared to be a significant relationship between past need for information on certain topics, and the type of organisation in which the questionnaires were distributed. For example, a greater proportion of respondents in CABx and other advice agencies had required welfare benefit and housing information, than had public library users.

When asked why information had been sought, almost half of the survey sample (48.8%) identified educational and study reasons, and a significant number also cited work-related reasons, family and personal reasons, and general interest. There was a fairly low incidence of information seeking for political decision making and for religious reasons.

Respondents were asked where they had gone to obtain information. The majority (66.7%) had visited a public library, with CABx forming the only other significant source (17.7%). However, not unexpectedly, a clear relationship emerged between the organisations cited and the type of organisation in which the questionnaires had been distributed. 88% of the respondents in public libraries cited public libraries as their source of information; 72.8% of CABx respondents cited Bureaux as their resource; while 84.3% of the users in the other advice agencies indicated that these agencies had been their preferred option.

Some interesting trends emerged when examining where respondents had gone to obtain particular types of information. Of the five most sought topics, public libraries were the most popular source for: educational information (84.4%), leisure and recreation information (95.5%) and information on health care (82.3%). However, for legal information, only 37.5% had gone to a public library while 50% had gone to a CABx or another advice agency; and for information on welfare benefits only 5% had gone to public libraries compared with 75% having gone to CABx and other agencies.

Respondents were also asked how satisfied they were with the information they obtained. Greatest dissatisfaction was registered by users of legal and welfare benefits information and by those that had sought information from Government departments.

196 (25.5%) respondents had encountered difficulties in the past in obtaining information. When asked to provide details, the majority indicated the types of information

sought had proved difficult to acquire, with information on legal issues, education, welfare benefits, health care and employment being the most prominent. A number of respondents to this question indicated why the information had been difficult to obtain. The most common reason was that relevant resources were inadequate or were unavailable. Some respondents had encountered difficulties with using various elements of information systems, such as catalogues, classification schemes, and electronic databases; others felt that official secrecy had prevented their accessing material or had found information full of jargon.

Using the same list of 23 subjects discussed above, respondents were asked if they felt they might want to find out more about any of these topics in the future. 1014 (78.4%) of the respondents predicted a future need for information. The top six subjects required in the past (leisure and recreation, education, employment, transport and travel, legal issues, and health care) were still regarded as the six types of information most likely to be required in the future, although in a slightly different order of preference. While there was a slight increase in the predicted need for legal information and health care information, less interest was anticipated in information on employment, transport, leisure and most significantly education. There was a significant increased perception of information need about the European Union in the future.

Women again had less interest in technology and politics, but a greater interest in family/personal information, health care, employment and education. For those aged 44 and under, predicted need was greater than past need for most subjects on the list; while for those aged 45 and over, predicted future use of the majority of subjects was less than past use. Students, jobseekers, and those respondents running a home displayed an increased future interest, compared to past need, in the majority of the topics; while predicted interest from retired respondents was lower in the majority of subjects. There were also a number of geographic variances, the causes of which were again not immediately obvious: for example, future interest in equal rights ranged from 6.1% in the East Midlands to 24.7% in the North West. Respondents from ethnic minority groups, when compared with white respondents, again displayed a preference for information on immigration and equal rights. Disabled respondents showed a greater interest in welfare benefits and equal rights, but less of an interest in employment and politics.

The four most popular past reasons for seeking information (educational, work-related, family, and satisfying a general interest) are also regarded as the most likely reasons for requiring information in the future. There was, however, a significant reduction in predicted educational reasons; it might be hypothesised that this relates to difficulties in predicting future areas of educational deficiency. Political decision making and religious reasons were again regarded as the least likely motives for information seeking.

Respondents were given a list of organisations and people and asked to indicate whether they would approach them for information. In total, 1209 (93.4%) of the respondents indicated they would approach at least one of the sources to obtain information.

*Table 2. Preferred sources of information.*

| Organisations / People | No. | % |
|---|---|---|
| 1. Public libraries | 1001 | 77.3 |
| 2. Family and friends | 798 | 61.7 |
| 3. Offices of Government departments and agencies (e.g. Inland Revenue, Benefits Agency) | 697 | 53.9 |
| 4. Post Offices | 687 | 53.1 |
| 5. Citizens Advice Bureaux | 652 | 50.3 |
| 6. Professional people (e.g. doctors and social workers) | 634 | 49.0 |
| 7. Local council offices | 551 | 42.5 |
| 8. Academic libraries | 509 | 39.4 |
| 9. Other information and advice centres | 406 | 31.4 |
| 10. MPs | 331 | 25.6 |
| 11. Professional / Trade Associations | 293 | 22.6 |
| 12. Chambers of Commerce | 121 | 9.4 |

This is felt to be a significant area in terms of overall consideration of sources of access to information used by the public. Over three quarters of respondents would use public libraries. Between half and three quarters of the respondents would approach Citizens Advice Bureaux, Post Offices, Government departments and agencies, or their family and friends. Additional preferred sources of information included: the Internet, leisure centres, local councillors, interest groups, banks, community centres, trade and funding agencies, PTAs, religious organisations, schools, telephone help lines, tourist information centres and voluntary agencies.

Women displayed a significantly greater preference for using five sources of information: family and friends, professional people, advice agencies, Post Offices and CABx. For Post Offices, public libraries, academic libraries, advice agencies, Chambers of Commerce, professional people, and family and friends, the trend was for interest in using these sources to wane as respondents grew older. Use of Government departments ranged from 63.2% in the South East to 40.4% in Yorkshire and the Humber; while use of Chambers of Commerce ranged from 18.8% in the North West to 4.9% in Scotland. In all regions public libraries were the most frequently cited resource, from 68.8% of respondents in Yorkshire and the Humber to 93.2% in Wales. Respondents in rural areas, compared with those in other areas, displayed a significantly greater interest in obtaining information from local council offices, professional people, Post Offices and public libraries.

Ethnic minority groups displayed a significant preference for visiting advice agencies, academic libraries, Post Offices and Chambers of Commerce; although public libraries (77%), and family and friends (65%) remained the two most favoured sources. The resource most frequently cited by disabled respondents was CABx (58.5%), and when compared with those without a disability, a significant preference was also shown for contact with MPs. Disabled respondents felt less inclined to obtain information from public libraries, academic libraries, family and friends and Post Offices.

Significant differences in preferred sources of information were also identified when examining the type of organisation in which the questionnaires were disseminated. For the users of CABx and advice agencies, CABx were their preferred source (78.7% and 56.1% of respondents, respectively), followed by family and friends. While public libraries were third on the list of CABx users' preferences, they were actually sixth on the advice agency users' list, behind Post Offices, Government departments and professional people.

The survey sample was asked in which language respondents prefer to obtain information. Excluding the Welsh and Irish languages, only 1.4% indicated a preferred minority language, compared with 7.7% of respondents that came from an ethnic minority.

From a list of methods of obtaining information, respondents were asked to indicate their favourite three methods in order of preference.

*Table 3. Preferred methods of obtaining information – ranked by weighted scores.**

| Method | Points |
|---|---|
| 1. Talking face to face with someone | 1023 |
| 2. Reading a book | 848 |
| 3. Looking through a collection without help from the staff | 827 |
| 4. Reading a newspaper | 679 |
| 5. Talking by telephone to someone | 518 |
| 6. Listening to the radio | 426 |
| 7. Watching television | 411 |
| 8. Reading a leaflet / pamphlet | 312 |
| 9. Using a computer | 275 |
| 10. Reading a magazine | 232 |
| 11. Writing a letter | 171 |

* Where three points were awarded for a 1st choice, two points for a 2nd choice, and one point for a third choice.

These data suggest that no single mechanism for enabling access to information should be seen as the ultimate solution to the information needs of the citizen. Rather a complementary range of solutions must be offered to the citizen.

There were a number of differences between age groups: use of a computer was cited as a preferred method by 40.6% of those aged 19 and under, and by 22.1% of those aged 20-29, but by just 10.6% of those aged 30 or over. The use of computers was cited by 33% of students, but just 9.7% of those running a home and 3.4% of retired respondents. Regional variations were prevalent, but once again no clear trends were evident. Acquiring information through face-to-face communication, for example, was preferred by 60.2% of

Scottish respondents, but only 30.6% of those in London. Ethnic minority respondents indicated a preference for obtaining information by using a computer and watching television. Disabled respondents, meanwhile, displayed a greater preference for listening to the radio and face to face communication.

Significant variances were also found when comparing preferred methods with the type of organisation from which the respondent obtained the survey form. For example, reading a book was preferred by 48.4% of public library users compared to 12.1% of the users of CABx and advice agencies, while reading a magazine was cited by 16% of library users compared with just 7.4% of Bureaux and agency users. Face to face communication, on the other hand, was preferred by 68.8% of Bureaux and agency users compared with 42.2% of library users; while telephone contact was cited by 40.3% of CABx and agency users, but only 24.7% of library users.

Respondents were asked how often they would use computers to look for information, if public access to computers was made more widely available. In all, 969 (74.9%) of the respondents indicated they would use computers on at least an occasional basis in at least one of the places listed.

*Table 4. Use of public access computers on at least an occasional basis.*

| Location | No. | % |
|---|---|---|
| Computers in public libraries | 940 | 72.6 |
| Computers in Post Offices | 488 | 37.8 |
| Computers in shopping centres | 454 | 35.1 |
| Computers in town halls | 378 | 29.2 |

There was a clear emphasis on public libraries as an appropriate location for computerised access to information for a significant majority of the general public.

In the final question, respondents were asked if they believed that access to accurate and unbiased information is important for exercising their rights as citizens. A highly significant majority (79.2%) believe that access to high quality information is very important for exercising their rights as citizens.

It was significant that younger respondents and students felt less certain of the importance of information than the other age groups. Again there were inexplicable regional differences. Belief that information was very important ranged from 90.6% in the South East to 67% in Yorkshire and the Humber.

## Conclusions

It is felt that the survey response (1294) was sufficiently representative of the UK population for generalisations to be drawn. The majority of respondents (59.4%) had sought information, while just under a quarter (23.8%) had experienced difficulties in accessing information in the past. Information need related to a range of categories reflecting the significant areas of citizens' experience. Educational reasons for information seeking were the most frequently cited both in the past and the predicted future. However, other reasons, such as work-related and a general interest, were significant.

Over three quarters (77.3%) of respondents said they would use public libraries on at least an occasional basis. Between 50% and 75% of respondents would approach Citizens Advice Bureaux, Post Offices, Government departments and agencies, or their family and friends.

Face-to-face communications were preferred by the largest number of respondents: however, the traditional book retained its appeal with the second highest number of citations. Other preferred options included browsing in a collection and by telephone. The media, too, were popular, with newspapers being more frequently cited than TV or radio. Only a small proportion of respondents expressed a preference for using a computer to seek information.

These data suggest that no single mechanism for enabling access to information should be seen as the ultimate solution to the information needs of the citizen. Rather a complementary range of solutions must be offered to the citizen.

There was a clear emphasis on public libraries as an appropriate location for computerised access to information for a significant majority (72.6%) of the general public. However, other public places, such as Post Offices, shopping centres and town halls, would attract a significant body and proportion of the general public.

A highly significant majority (79.2%) of the respondents believe that access to accurate and unbiased information is very important for exercising their rights as citizens.

## References

1. Marshall, T.H. *Citizenship and social class.* (Cambridge: Cambridge University Press, 1950.)

2. National Consumer Council. *The fourth right of citizenship: a review of local advice centres.* (London: National Consumer Council, 1977.)

3. Policy Studies Institute. *Information and citizenship in the United Kingdom.* Draft report prepared as part of the European Commission's Information and citizenship in Europe study. (London: Policy Studies Institute, 1995.)

4. Library Association. *Community information: what libraries can do.* (London: Library Association, 1980.)

5. Donohoe, J.C. 'Community information services – a proposed definition.' In Martin, S.K. (ed.) *Community information politics. Proceedings 39th ASIS annual meeting* 13. (Washington: American Society for Information Science, 1976.)

6. Bruce, I., McKennel, A. and Walker, E. *Blind and partially sighted adults in Britain: the RNIB survey.* (London: RNIB, 1991.)

7. Tinker, A., McCreadie, C. and Salvage, A. *The information needs of elderly people: an exploratory study.* (London: Age Concern Institute of Gerontology, 1993.)

8. Warner, E.S., Murray, A.D. and Palmour, V.E. *Information needs of urban residents.* (Washington, DC: US Department of Health, Education and Welfare, Office of Education, Bureau of Libraries and Learning Resources, 1973.)

9. Beal, C. 'Studying the public's information needs.' *Journal of Librarianship* 11(2), 1979, pp.130-151.

10. Library Association. *Library Manifesto.* (London: Library Association, 1997.)

11. Library and Information Commission. *Review of the public library service in England and Wales for the Department of National Heritage: final report.* (London: Aslib, 1995.)

12. Usherwood, R. *The public library as public knowledge.* (London: Library Association, 1989.)

13. Swash, G. and Marsland, G. 'Current trends in the use of the public library information centre.' *Library Management* 15(6), 1994, pp.20-22.

14. Great Britain: House of Commons. *Code of practice on access to government information.* (London: HMSO, 1994.)

15. Great Britain: House of Commons. *Your right to know: the Government's proposals for a Freedom of Information Act.* (London: The Stationery Office, 1997.) (Cm 3818)

16. Great Britain: House of Commons. *Crown copyright in the information age: a consultation document on access to public sector information.* (London: The Stationery Office, 1998.) (Cm 3819).

17. *government.direct: a prospectus for the electronic delivery of government services.* Available at: *http://www.open.gov.uk/citu/gdirect/* (visited March 1997)

18. Government Information Service home page. Available at: *http://www.open.gov.uk* (visited June 1998)

19. UKCOD home page. Available at: *http://www.democracy.org/uk/* (visited February 1998)

20. Library and Information Commission. *New library: the people's network.* (London: LIC, 1997.) Available at: *http://www.ukoln.ac.uk/services/lic/newlibrary/full.html*

21. Parliamentary Office of Science and Technology. *Electronic government: information technologies and the citizen.* (London: Parliamentary Office of Science and Technology, 1998.)

# The migration of news to the web

*Peter Williams and David Nicholas*

*Internet Studies Research Group, Department of Information Science, City University, London EC1 0HB.*

*Email: pw@soi.city.ac.uk*

**Abstract** *Virtually all UK and US newspapers and the vast majority of regional and even local titles are now represented on the web. Indeed, the Yahoo news and media directory lists no less than 114 UK newspapers online (as of November 1998). Broadcasters from the BBC and Sky downwards, and all the famous news agencies (Press Association, Reuters etc.) also boast comprehensive Internet services. With such an array of sources available, the future of mass access to the Internet, possibly via TV terminals, suggests that more and more people may soon opt for this medium to receive the bulk of their news information. This paper gives an overview of the characteristics of the medium illustrated with examples of how these are being used to both facilitate and enhance the content and dissemination of the news product. These characteristics include hyperlinking to external information sources, providing archive access to past reports, reader interactivity and other features not possible to incorporate into more passive media such as the hardcopy newspaper. From a survey of UK and US news providers it is clear that American newspapers are exploiting the advantages of web information dissemination to a far greater extent than their British counterparts, with the notable exception of The Electronic Telegraph. UK broadcasters, however, generally appear to have adapted better to the new medium, with the BBC rivaling CNN in its depth and extent of news coverage, use of links and other elements.*

## History – newspapers online

The current explosion of the Internet, talk of the 'dawn of the information age', and consequent interest in web-based news sources, have obscured the fact that electronic news provision has been around for almost 30 years. In fact the first experiment in disseminating news electronically began in 1971 when the General Post Office in the UK began work on what was to become the Prestel service. Launched in 1979 it offered news bulletins, home banking, flight bookings and other information via special TV-like monitors. However, high costs and various other factors meant that by 1986 only 65,000 people had subscribed to the service – many of them travel agents, for whom the benefits made the expenditure worthwhile. The project was abandoned in 1993, by which time usage had dropped to 30,000[1].

Ceefax also arrived during the 1970s. Unlike Prestel, however, it has continued to survive and, indeed, flourish. It was developed by the BBC out of research into subtitling programmes for the deaf. The fall in the price of computers, development of ROM technology, and advances in data transmission techniques all helped the project, and it soon became clear that applications beyond the basic service initially envisioned were possible[2]. Test pages were operational by 1973 and trial broadcasts begun in 1974 before a national launch in 1976. Oracle, the ITV version of teletext, was developed in tandem, with television set manufacturers also being consulted at each step. The power to transmit live news in (tele)text form was realised during this testing. This is highlighted by Sigel[2], who describes how, at match point in the 1975 men's final at Wimbledon, the Ceefax sub-editors, watching the game on TV, prepared two separate news flashes. One read '5:5 in the fourth set, with Ashe leading by two sets to one' and the other 'Ashe wins Wimbledon 6:1 6:1, 5:7, 6:4'. He claimed that the correct caption, (which was the final result) had been broadcast 'before the umpire had had time to pronounce 'game, set and match"!

Adoption of teletext services were slow in the beginning, with only 3% of households having access up to December 1981. Growth became more pronounced in the 1980s, with take up at 17% by 1987[3]. Now, however, teletext broadcasts to a mass audience: 60% of households has at least one Teletext TV. The service is used daily by 9.4 million people and weekly by 19.2 million, according to research by polling organisations NOP and Kaleidoscope[4].

Meanwhile, in the US, Knight-Ridder were experimenting with another electronic news system 'Viewtron', tested in Miami in 1981. This initiative, like Prestel, was also unsuccessful. Cheap cable television, becoming widely available at the time, inhibited sales of the Viewtron terminals, and although the company persisted for a number of years the system was finally abandoned in 1986, after incurring substantial losses for the duration of the experiment.

The French, however, were more successful with their Minitel electronic information and telephone directory system which began operating in 1981. According to Beckett[1], there were two major reasons for this. Firstly, the terminals were given away free by the then socialist government. Secondly, the service gave people what they wanted and needed – the telephone enquiry system which it had been designed to replace was notoriously inefficient. Millions of people around the country signed up for the service. Interestingly, Minitel probably provided the first example of a hardcopy newspaper being scooped by its electronic form, when *Libération* used the medium to publish results from the Los Angeles Olympics well before the first edition of the paper appeared on the streets.

In a parallel development the potential for accessing news online was spotted by the – then embryonic – information industry. Beginning with *The Guardian*, in 1985 World Reporter (now FT Profile) mounted a campaign which was to see the (near) full-text of all broadsheets, most tabloids and a good number of regional newspapers on the service by 1997. The cross file searching facility UKNEWS offered the prospect of searching several dozen British newspapers at the same time and viewing a composite set of results. Also in a ground-breaking event the *Financial Times* appeared online on FT Profile before it hit the newsstands.

Possibly encouraged by Minitel's success IBM and Sears developed a news and communications network called Prodigy. The launch of this, in 1987, coincided with the advent of the affordable home PC and expanding interconnectivity between computers, especially at academic and research institutions. As Beckett[1] puts it 'electronic newspapers could now piggyback on existing computer networks with lots of technically literate, even adventurous users, rather than having to start from scratch'. Other services began to spring up to capitalise on the new environment, with Knight-Ridder, despite having had it fingers previously burnt, among them. It joined forces with America Online, a subscriber network inspired by Prodigy, and started to develop the Mercury Center from its Information Design Laboratory in Boulder, Colorado.

Some of the most significant steps in the development of online newspapers have come from The Mercury Center. The local paper, *The San Jose Mercury News*, went online in 1993 and has continued to be in the forefront of electronic news publishing. The paper pioneered 'extra' services such as a news archive dating back to 1985, expanded local information, and an electronic Bulletin Board for readers to communicate with each other and with staff. Similarly innovative, the paper also incorporates a personalised news service in which readers enter keywords, weighted by the user in terms of degree of relevance, and receive articles dispatched via email[5].

*The Electronic Telegraph*, the online version of *The Daily Telegraph*, was the web newspaper pioneer in Britain, with its first Internet edition appearing in November 1994. *The Times* had launched '*ST:Online*' in the September of that year, which included an interactive discussion forum, but this was a modest text only service not incorporating the then very new web technology, although it did provide a gateway to the Internet. The full web editions of *The Times* and *Sunday Times* were launched on 1 January 1996, and were the first UK papers to include (virtually) the full text of the hardcopy versions. *The Electronic*

*Telegraph* had adopted early a policy of editing or reshaping stories for the online reader – although nowadays the online paper, if not carrying the full hardcopy text, actually includes far more material than its hardcopy counterpart, i.e. carrying web only copy.

By the middle of its first month of operation *The Times* was already claiming that 2000 readers per day were registering to access the service, and that 'in California, *The Times* has become the essential afternoon paper'[6]. Just six months later the online paper won the 'hottest site on the Net' award at the first UK Web Awards ceremony held in London at the British Academy of Film and Television Arts (BAFTA), sponsored by Yellow Pages online and *.net* magazine. The Internet editions of *The Times* and *The Sunday Times* became the first national newspapers to receive an audit certificate from ABC//Electronic in August 1997[7]. *The Guardian* was typically late going online, initially developing a number of discreet pages reflecting readers' interests. These included a Webzine for science and technology and specialist pages looking at the Scott report and MPs' register of interests. It made its mark with the Euro96 site, which obtained a claimed 500,000 hits a day, and 'was one of the 100 hottest sites in the world'[8]. As is discussed later in this paper, current research at City University is casting doubt on the validity of inferring 'hits' statistics into real users or even degree of use.

## Elements of an online newspaper

Many news providers have already taken steps to enhance or add value to their printed product by exploiting the potential of the Internet. The technology allows for enhanced and improved information provision not possible in a printed, or, indeed, passive broadcast medium. Thus web papers offer the potential for:

- comparatively unlimited space availability – permitting greater depth/coverage of news stories, the provision of large archives and the mounting of added value reference services;
- currency/immediacy;
- hyperlinking;
- multimedia;
- customisation;
- interactivity.

## Space availability

The sheer 'space' afforded by the electronic environment is a major factor in the design and content of online newspapers. During a recent study of journalists' use of the Internet[9] one reporter claimed that producing a daily newspaper was like squeezing 'a swimming pool of information' into 'the same quart pot every day'. With an electronic carrier that problem disappears at a stroke – space is only constrained by the amount of RAM on the server hard discs. This means that literally thousands of pages may be posted every day without the worry of running out of space. The benefits are threefold. Firstly stories can be provided to add greater depth or background; secondly, there is room to include items that have lost out in the daily battle for hardcopy space (a newspaper produces far more stories than it can realistically publish); and finally, material can be archived on site, and those archives made publicly accessible.

Newspapers have not been slow to recognise these possibilities, even if in the beginning editors were content to limit web material to some text lifted from the paper and a masthead. Nowadays online editions invariably include a 'value added' element, with seemingly every paper offering, in one respect or another, a more comprehensive product than would be possible in hardcopy. American titles have gone furthest down the road, with *The Wall Street Journal*[10] claiming to have 100,000 pages available for perusal at any one time – although the site is only accessible by (a modest) subscription.

### *More comprehensive coverage of current news items*

One of the best examples of this comes from *The San Jose Mercury News*' series on the involvement of the CIA in heroin sales. It augmented its investigative reports with biographies of key players in the story, court filings, texts of CIA teletype memos, a photograph of an arrest warrant, and even articles reproduced from other papers that were critical of its reporting[11]. The Kosovo crisis was at its height during the preparation of this paper – online coverage has been greatly enhanced by the big US newspapers' web coverage, with links back to previous articles, full text reproductions of important documents and links to external news and information sources.

In Britain, although online papers are generally not quite so comprehensive, they are catching up fast. *The Electronic Telegraph*, for example, incorporates a 'Euronet' section in its online publication, featuring a timeline of events surrounding the development of the European Union, essays by politicians, business people and academics, readers' polls and fora, a 'handy' guide to EU history and a collection of key documents such as The Maastricht Treaty. BBC Online includes special reports and background information of top stories. Taking the topical example of the crisis in Kosovo (current at the time of writing) the BBC web site included a vast array of background details to accompany and supplement its current news coverage. The most innovative of these was an 'interactive map' of the region. Readers are invited to click on any particular region to call up, on a side frame, a brief account of the involvement of the various military and political factions, ethnic groupings etc. within that area. Another page is titled 'History file: Yugoslavia and the Balkans'. Included here is a static map with a timeline running underneath in the form of a series of calendar icons showing key years in the history of the troubled region. Activating an icon calls up an account of important events of the time. Elsewhere are postings which variously profile Slobodan Milosevic and Ibrahim Rugova, analyse NATO's military options and outline world opinion on the situation. The site also itemises other stories from recent days dealing specifically with the (at the time) latest developments. This means that, conveniently, a site search is only required for earlier material.

### *Archive facilities*

Considering that so much news text is already in electronic form[12] surprisingly little has been done by the UK press to capitalise on the opportunities. *The Guardian* provides free access to its Online section electronic archives (ironically, the section that dates most quickly), going back to 1995, but has no facility to search other sections, although this is promised in the future. All stories that have appeared on *The Electronic Herald's* web site are archived and accessible via a search engine. This service is considered as one of the highlights of the site. Online editor Teddy Jamieson is convinced it will be 'a resource tool for academics, politicians and the general public by providing an easily accessible, authoritative source of information'. Indeed, the service is seen to offer the possibility of building a new relationship between reader and paper[13]. At present the paper uses a search engine provided by Excite, which supports both free text and Boolean commands, but Jamieson claims that in the future users will be able to specify dates, section of the paper or use other criteria in their searching.

The News International buildings at Wapping houses five computers that contain more than 50 gigabytes of hard disk capacity and two gigabytes of computer memory – enough to house the complete Internet editions of *The Times* and *Sunday Times* for at least the next ten years[14]. News International is exploiting this datastore. *The Times* allows searches for back copies of its paper from the day it first went online, on 1 January 1996, and has developed an 'InfoTimes' section, where important articles and long-lasting information are grouped under headings. Retrieval is by a directory system rather than by word searching. As with other aspects of web news dissemination, BBC News online is far ahead of most newspapers, if not *The Electronic Telegraph*. Again, a comprehensive archive is on the site, searchable by keywords. The full text as it originally appeared on the web site is accessible, usefully including the links to external sources made at the time.

In the US archive access is much more common. *The San Jose Mercury News*, mentioned earlier, has a 'Digital News Library' on its site, housing more than a million articles. These include not only all the *Mercury's* own stories since 1985, but also the archives of 19 other papers hosted by Knight-Ridder. Searchers can access headlines and the first lines of stories for free, and then pay around $1.95 to be sent the full text. This sum, users are advised, can vary slightly depending on the source. *The Washington Post* has an excellent archive service – articles from previous editions of the paper can be accessed free for content less than 14 days old, and $1.50 or $2.95 (depending on time of day) for older material. Impressively, the database dates back to September 1986.

***Reference services***

Many papers are providing comprehensive reference services in their web editions. Perhaps the best example of this is *The New York Times*' education supplement. This includes resources on a wide variety of subjects to aid teaching – right down to the detail of lesson plans themselves; school calendars, and other information for students and teachers. Special features on various topics are also offered. On 24 October 1998, for example, themes included 'The launch of Sputnik' and 'Little Rock remembered'. US newspapers, being regional, also tend to have excellent local and regional information on their sites. For example *The San Jose Mercury News* includes a 'Just Go.com' section, with voluminous listings of what's on in the local (Bay) area covered by the hardcopy paper, and a customised directions facility detailing how to get to the event of your choice from your own home[15]. *The Washington Post* and *New York Times* are other papers with comprehensive guides to restaurants, cinema listings and events calendars for their respective cities.

In the UK, *The Times* 'InfoTimes' section provides a general reference service for readers on such subjects as stress management, personal finance and computing for small firms, plus a number of sports facts and figures. It also incorporates an extensive student supplement, containing study guides, links to universities and a full information service. Two 'mini-sites' currently being created by *The Electronic Herald* are dedicated to Scottish devolution and Glasgow's year of architecture and design respectively. These initiatives not only include material exclusive to the Internet, but are also linked to the paper's extensive archives, allowing readers to peruse articles published in the paper on both topics in recent years[13]. An idea adopted by *The Guardian* is to celebrate the millennium by plotting the history of the last 1,000 years in 500 days. The online paper provides a complete reference guide, in two year blocks, to events and characters that have shaped this period of history. Hyperlinks to external sources (such as 'The Viking Network' etc.) are highlighted in the text. The hardcopy version of the paper runs a brief summary in a small column everyday in '*G2*', the magazine section, where the full Internet version is prominently advertised.

Often reference pages are predominantly or even exclusively composed of links. These are discussed in the next section.

**Hyperlinking**

The ability to jump seemlessly from one electronic file to another, be it on the same computer, the same LAN or on the other side of the world, is the outstanding (siren) feature of the World Wide Web. For commercial news organisations, however, this characteristic of the system is not necessarily wholly beneficial. Hyperlinks are the backbone of the web – indeed, the very name 'web' implies this interconnectivity. On the other hand, news providers are rightly wary that those jumping from their site to another may not come back again, and for this reason much of the national UK press has, up to now, been reluctant to utilise this aspect of the htm web language. Many papers – *The Independent* and *The Times* are examples – do not link any of their stories to external sites, although the latter does use linking in its reference pages. Others, however, are more adventurous, with *The Electronic Telegraph* leading the way in including links to appropriate alternative information sources for many of its online stories and as an integral part of its reference pages.

Again, US news sources appear more liberated and offer a more comprehensive service. *The Washington Post* linked its Kosovo reports to an Associated Press photojournalist's photographs and diary notes from Kosovo, Office of the High Representative in Bosnia and Herzegovina, NATO, BosniaLINK, the US Department of Defense information service, the UN High Commissioner for Refugees, and others. Although impressive, and certainly providing an almost limitless amount of information on the area, there is, nevertheless, a certain omission in that none of the links are to interested parties on the ground – all lead to either US or international information providers. This is in contrast to CNNi links, which include the Albanian Ministry of Foreign Affairs and, representing the Serbia point of view, the official site of the Government of Yugoslavia. However, it is in linkage to documents which gives *The Post* the edge even over the impressive CNNi coverage. The documents referred to are the full texts of The Dayton Peace Agreement, and UN Security Council Resolutions.

No doubt this extra information provision (or facilitation) gives host sites prestige, and a reputation for being good information sources, in addition to their prowess in reporting events themselves. If some elements of the UK national press have not grasped this nettle, it appears that their regional competitors certainly have. The approach of one such newspaper to the blessings/evils of hyperlinks has been documented elsewhere[16]. Suffice to say here that an editorial decision was made that the paper in question (*The Northern Echo*) would attract readers by the provision of an information service that included primarily links to external sites and organisations, and that this attraction would outweigh the dangers of readers jumping off. Those doing so would come back to *The Echo* site when the next information need arose. These links included, courageously, routes to many rival UK and US national and regional papers.

Web sites made up exclusively of active hyperlinks are currently termed 'portal sites'. These are dedicated, as the name implies, to providing a gateway to other sites on a particular subject or subjects. It appears that the regional press in the UK are far more enthusiastic about providing this kind of service than the nationals. Apart from the reference service of *The Northern Echo*, Eastern Counties Network, the parent company of newspapers in the East Anglia region, incorporates links to business, tourist and other sites on its excellent web pages. Of the national press, *The Electronic Telegraph* again leads the way. Its 'Euronet' pages include a resource section providing links to locations containing key policy/legislative documents and other relevant external sites dealing with the European Union.

In the wider context of news provision on the web, major Internet search directories such as Yahoo! contain long hyperlink lists of newspaper and other news sites, and The Electronic Newsstand provides links to literally hundreds of online magazines from a variety of countries and in many languages. A more focused portal site is that of Boston.com. Apart from being a source of information generally about Boston – arts, weather, attractions etc. – the site also offers access to the content of no less than 18 local newspapers and broadcasters. These include *The Boston Globe* and WGBH, Boston's famous public broadcaster.

## Currency

Of course currency is *the* key defining element of what constitutes news. And it is plainly what provides the online dissemination of news with a huge in-built advantage. Thus online newspapers are freed from the constraints of print production deadlines and distribution timetables. For commentators like Robin Hunt, ex-New Media Editor of *The Guardian*, it also provides the principal reason why all breaking news will migrate to the web, leaving the hardcopy newspaper as a features driven product. *Libération's* dissemination of sports results electronically hours before it went to print has been mentioned, though this may not have been the first electronic 'scoop'. There are conflicting claims for the first Internet newspaper to 'scoop' itself on the web. Hanson[17] regards *The Dallas Morning News* to be the first to undertake this practice, in February 1997. The paper was in possession of the information that Oklahoma bombing defendant Timothy McVeigh had allegedly confessed to his lawyers that he had, indeed, committed the criminal act. Staff on the paper were worried that, once they began seeking reaction to the news to include in their story, there was a possibility that the defendant's lawyers might break the news themselves, in order to put their own interpretation on it and to pre-empt the press. The solution for the paper was to post the story on the Net. Jamieson, however, claims that his own paper, *The (Glasgow)*

*Herald* reported the earlier Dunblaine tragedy on its web site 12 hours before the first print story appeared. With regard to the non-English language press, Nicholas, Frossling and Buesing[18] cite the Swedish paper *Dagens Nyheter* which was regularly offering 'front page news' the night before its print release back in 1996.

Richard Withey[19], New Media Director, News International, says that there is still a huge reluctance amongst editors and journalists to break a story over the web. Breaking a story is the most important thing a newspaper can do and no journalist wants to hand that baton to the web paper (yet).

As yet newspapers have not taken up the option of continually updating stories to any great extent, though it is not clear whether the pitfalls outlined above have made them wary, or if they simply lack the staff or conditions to modify their sites as needed. *The Times* adopts a very simple but effective approach. The paper retains the same 'front page' throughout the day, but includes a 'news update' option on its top menu bar. This is an area where breaking news is posted, usually in brief one or two paragraph messages without comment or analysis. No similar feature was found on any other newspaper site. On 8 October 1998, for example, the 'front' pages of *The Guardian, Electronic Telegraph, Independent*, and *Mirror* were all accessed at hourly intervals between 10am and 4pm to monitor web site news updating, the content of each one remaining the same[i].

Newspapers in the US appear to give more priority to currency. Many newspapers state the time articles were posted, and the front page may change more than once during the course of a day. *The Washington Post*, however, in a similar manner to *The Times* in England, has also managed to update stories and still mirror its hardcopy twin by providing a facsimile of the paper's front-page. Unlike *The Times*, however, it does alter the main body of its web edition to reflect breaking news.

Broadcasters are used to dealing with news as a continuous stream, and the BBC has had no trouble in offering a rolling news service to readers, adding and modifying pages round the clock. This is not just for developing news stories – the comment column article by Rev. Doug Gay that caused his dismissal from News International was published on the BBC site at 18:11 UK time, and a 'talking point' article on unemployment (where readers are invited to post their comments on particular issues) was initiated at 14:46. CNN adopts a similar approach. Both channels thus approximate their original roles as broadcasters, while the newspapers have, up to now, tended to produce web 'editions' of their product, in some cases simply reproducing copy from their paper versions (or, as Katz[20] put it, 'shovelling their content on to the web') enhancing these with reference and archive services.

Hanson[17] cautions against a rush to get information into the public domain as quickly as possible. Apart from what he describes as an emphasis on 'getting facts on to the screen immediately, with little patience for enterprise and investigation' he cites three main reasons why instant news may not be as desirable as Internet enthusiasts are assuming. These can be listed as follows.

### *'Stenography for the status quo'*

This is the danger that online papers will report what officials, politicians and others say without analysis or assessment of the validity of the statements. Under this heading too is the similar problem that the same officials etc. will find it easier to set the news agenda. This is rather ironic in the light of the claimed ability of the medium to allow news consumers to set their own agendas.

### *Looser standards*

The worry here is that there will be an erosion of the normal standards of accuracy. Hume[21] is also concerned about this problem. She noted that the scramble for 'exclusives' by the local press and broadcasters during the O.J. Simpson trial backfired for two reasons. Firstly, people questioned about the coverage had no idea of who released what news story in the saga first, nor were they particularly concerned. Secondly, the pressures on staff to get information out were so great that much of the coverage was factually inaccurate. Plainly a case of the dumbing down of news coverage.

***A focus on 'the flavour of the minute'***

This is the potential problem of having to provide a constant novelty value, as the news channel is open 24 hours a day. Hanson[17] highlights news providers such as CNN who 'must devote vast amounts of effort to satisfying the public's appetite for novelty for just the next half hour'. He goes on to add that 'a newspaper that commits itself to the online world is likely to find its resources dissipated in just such a shallow pond, with little energy left over for the good writing, clear explanation, in-depth investigation ... that mark the best newspaper journalism'.

**Customisation**

This is where the reader has some control over the information received from a news organisation. At a simple level it may consist of a service in which readers are invited to request only certain sections of each day's output – be it sport, politics, or news about a particular country etc. – which is delivered by email. A more sophisticated system, such as that offered by CNN, involves creating a detailed user profile from very specific category options, and prioritising them for a personalised clippings service. Another method of customisation is to create functions whereby the reader can enter personal details, and receive appropriate information according to stated circumstances. An example from MSNBC on the Internet (*www.msnbc.com*), described by Pavlik[22], illustrates the potential:

> 'On February 21, NBC's Dateline ran a piece about dangerous roads in America, zeroing in on three particularly treacherous thoroughfares. The program invited viewers to log onto the MSNBC site to learn about roads in their community. Those who did so could enter their zip code and, within seconds, based on federal data, find out how many fatal accidents had occurred in that community between 1992 and 1995 and on which roads. Within twelve hours MSNBC logged 68,000 visitors to that feature.'

Money Online provides another example. When presidential candidate Steve Forbes was championing his flat rate tax idea in 1994, a feature allowed people to key in their earnings profile and see how the proposed tax would affect them.

Not only traditional news providers are offering customised services, but search engines and directories too, although these are (probably) exclusively of the user keyword profile type. Yahoo!, for example, includes a 'My Yahoo!' selective dissemination service. Users register to receive 'personalised news' on their Yahoo browser page. There are several options including News and Current Events, Business and Investing, Technology, Entertainment, Arts and Culture, Sports, Lifestyles, etc. Possibly the most impressive example of customisation comes, fittingly, from one of the pioneering online newspapers, *The Raleigh News and Observer*. A feature of the site is the 'Nando News Watcher' which sends (or 'pushes') continuous feeds of customised local, regional and national/international news to a small window on the reader's screen, where it can be enlarged and viewed as desired.

Newspapers in the United Kingdom do not appear to have travelled far down this road yet, and there appears to be some opposition to doing so. MacArthur[23], for example, complains that what it produces is not really a newspaper at all but a piecemeal collection of articles without any coherent editorial focus. Patrick Ensor, Editor of *The Guardian Weekly*, is more enthusiastic. His paper has established a system whereby subscribers to the hardcopy version of the paper (principally ex-patriots receiving the paper by airmail) can register to receive an email version of the sections of their choice. This service clearly answers the complaints from readers who routinely have to wait weeks (five in the case of one oceanic subscriber) to have their paper versions delivered. These people – those on the Internet anyway – are now able to read current items such as sports results the day the paper is published or earlier. The service is considered an excellent compromise by the editor, who points out that readers get their news faster, but crucially are still required to pay for the full hardcopy edition. With many papers, including the daily *Guardian* itself, putting huge amounts of news (and investment) on the web, it is doubtful whether the payment system will be able to sustain itself for much longer.

ITN has just announced 'Britain's first personalised desktop news service', although, in fact, the only degree to which you can personalise it is by selecting certain categories of required news. As there are only three of these (World, UK and Business), the 'personal' nature of the service is perhaps a misnomer, notwithstanding the fact that the number of stories desired and the frequency of delivery can also be specified.

### Interactivity

Internet technology permits a far wider range of communication flow than that afforded by any other medium. Anything from private one-to-one emails to publicly broadcast newsgroup messages with a potential audience of millions are possible with the facility to reply to individual or all correspondents. In addition to this, the communication is (usually) almost instant, emphasising an element of interaction more dynamic than in written correspondence. This aspect of the Internet is one that has helped it gain appeal as a truly different and alternative phenomenon, and has led users to demand a degree of interactivity from the web that simply would not be possible elsewhere. Types of 'interactivity' may include bulletin boards hosting readers' fora on various subjects, where messages can be replied to individually or to the board; reader to journalist, where the authors of particular articles of interest make themselves available to answer emailed communication, and reader or journalist to public figure, in a similar public discussion.

Evidence as to the take up of interactive facilities is conflicting. Both *The Times* and *Electronic Telegraph* were already receiving up to 100 emails a day in 1996[24] and, according to staff at News International, *The Times* and *Sunday Times* are now (January 1999) receiving on average six times that figure[25]. Similarly, *The Guardian* letters editor John Shirley claims that as many readers' letters are received via email as through other media, with, interestingly, their quality 'on average' higher[9]. *The Herald* is one paper that plans a number of online readers' fora in the near future. This initiative is described by Jamieson[13] as being introduced so that readers can 'interact with the publication, to discuss the important issues facing Scotland' in a form of what is described as 'cyberdemocracy'. The enthusiasm with which *The Herald* is launching this service, however, may not be justified. Experimenting with online feedback at Newsquest Northeast, a major regional newspaper organisation, has proved disappointing. Only a handful of readers contributed to a cricket discussion forum, and not one message was received with respect to *The Northern Echo* leek growing competition[16,26].

Almost needless to say, across the Atlantic interactive services are well established, with papers regularly organising online debate – audio as well as written – between the public and local politicians etc., and as far back as 1996 Nicholas and Frossling[27] noted that any major event in the US, particularly disasters, tragedies etc. spawned a rash of message board sites and newsgroups discussing the situation, many of them hosted by newspaper sites, offering mutual support etc. Outing[28] regards this feature as one of the most important aspects of online news services.

Many issues are raised by consideration of potential, and actual, interactive practices in an online news environment. These include:

#### *The degree and nature of editorial staff contributions*

Despite much enthusiasm expressed in the US about journalists talking back to readers, answering questions, discussing issues etc., there is a seemingly insurmountable problem here of time availability. Nicholas *et al.*[9] found that in the UK at least, journalists were most reluctant to publicise their email addresses and, indeed, were reluctant to use the medium at all, for fear of being swamped by what they regarded as junk mail. And despite Alan Rushbridger's self proclaimed love affair with the Net[29], he – and many of his staff – seemingly almost on a point of principle fail to answer emails. A related issue is that of the maintenance of standards of accuracy and authority. Hanson[17] gives the example of *The*

*San Jose Mercury News*' 'already questionable' series on alleged CIA involvement in drug trafficking, being further discredited by Gary Webb, the reporter, participating in Internet 'chat room dialog'. The hapless journalist was said by Hanson to have 'made more sweeping claims than he ever had in the paper'.

### *Editorial moderation*

This is an issue coming to the attention of those with an interest in information policy and the law. Newspapers may be forgiven for being wary of letting anyone with any sort of obsession, complaint or general malevolence have their thoughts publicly accessible alongside serious professional news commentators. Indeed, Hunt[30] reports that *The Guardian* was forced to discontinue a memorial bulletin board in honour of Diana Princess of Wales due to the number of – to put it euphemistically – inappropriate messages received. Interestingly, however, it appears that online information distributors (such as newspapers) are not, according to conditions laid out in The Defamation Act 1996, liable for material posted on to sites they administer unless they play any part in deciding what is published[31]. A similar situation prevails in the US. In her paper on the development of the online edition of *The Washington Post* McAdams[32] describes how it was decided not to moderate communication, based on the fact that US case law stated that 'if an information service provider edits or censors discussion postings in any way, that provider may be open to [prosecution]'.

### *Newspaper accountability*

The implication here is that readers' greater access to the newsroom, their more proactive involvement in news analysis and their customisation of the news to fit their own personal needs may all pressurise publications into responding positively to demands and preferences thus expressed. Whether this will actually happen, and whether the process has already started, is debatable. Ian Katz, writing before his appointment as *The Guardian* New Media editor, is doubtful.

> 'Real power, as always, lies not in online exchanges but in daily story conferences among editors who don't read email. In fact, the familiar newspaper model lurks behind every icon: you can write us as many letters as you want, in a faster way than before, and we'll read them. But we're still going to decide what's important, and then we'll tell you'.[20]

On the other hand Derek Bishton of *The Electronic Telegraph*, claimed that such correspondence made the staff feel closer to the readers[24], although this may have been said bearing in mind his audience – the readers themselves!

### Multimedia

Although this is an important feature of the web and has led to the phenomenon of the multimedia journalist, practised and much beloved by Griffiths[33], exploitation of the potential of multimedia is currently minimal in online newspapers. This is hardly surprising – organisations traditionally geared towards producing text and pictures in paper format are unlikely to have the technical expertise or infrastructure to suddenly produce film journalism over night. Also, there is the question of need or desire. It is probable that newspapers do not feel they are compelled to fight broadcasters – at least, not on broadcasters' own terms. Indeed, it is difficult to imagine even national newspapers competing with the BBC to file video reports from overseas, or even accompanying the written word with sound files. The only paper surveyed for the current article seen to have even dabbled in additional media was *The Washington Post*, which broadcasts sound interviews with local government officials and other personalities.

However, there are indications that the future, despite the current low take up, may nevertheless lie along this path. The well known new media commentator Steve Outing[34] is not alone in claiming that the traditional newspaper will come under increasing financial pressure as 'the masses go online'. He goes on to add that 'news publishers will have trouble maintaining editorial quality as print revenues flatten out or ... decline'. his solution is for 'the smart media company' to diversify, 'taking advantage of the Internet to publish to multiple media and thus generate multiple revenue streams'.

Outing[35] describes just such an initiative happening in the US. *TCPalm.com* serves as the single news-oriented web site covering a four-county region in Florida. It operates with a staff of eight housed at *Stuart News*, one of six newspapers participating in the new project, although those developing it are not newspaper employees. They work with each of the papers – none of whom have a web site nor any new media employees – and a local NBC television affiliate, WPTV Channel 5, which provide content for the site. Thus it combines newspaper reporting with video from Channel 5 ported to the web, and is designed to be updated throughout the day. Indeed, for those with a fast enough Internet connection, the site can even replace TV news. Channel 5's newscasts are put on the web live, and can be viewed using RealVideo software.

Similarly, during a recent interactive publishing conference in Zurich, John Haile, Editor in Chief, of *The Orlando Sentinel*, described how his relatively small US newspaper is now a print, online, cable and radio channel medium. Journalists go out with notebook, camera, video and tape recorder; editing is done for all media at a style console[36].

## Online readership levels

The ultimate test of whether the news providers have been successful in the migration of news to the web is surely whether anybody has been using them. For many people the Internet is such a major part of their everyday working and/or leisure lives that it is hard to appreciate the as yet low penetration of the phenomenon, at least here in the UK. Indeed, even in the US, possibly the most 'wired' country in the world (with the possible exception, in per person terms, of the Scandinavian countries) well over half the adult population (62%) have never used the World Wide Web. Over here an astounding 85% are still web virgins. Thus, whatever value added elements are embodied into the sites of news providers' web pages, online readership levels are still drawn from a very low base.

Even within the online community, obtaining news from the web does not appear to be a priority. Despite figures produced by ZDNet[37] showing a decline in traditional newspaper and news weekly consumption attributable to the Internet[ii], another part of the study reveals conflicting evidence that using the web for news or current affairs information is not listed in the top 12 uses cited by respondents. In fact, the only 'news' like activities listed are finding sports scores, cited by 31% of respondents, and 'local information' – although this may well be of a general nature such as addresses of organisations, library opening times etc. Even among news reading web users, priorities were not with hard or breaking news. The Pew Research Center[38] found that the most popular type of news on the Net was the weather, with 64% of readers accessing such information. Neither finding is likely to offer much solace to investors in web newspapers.

Web audience figures provide more encouragement. ABC, the established audit company which monitors hardcopy sales, set up an electronic arm in 1996[39]. It regularly gives figures of over seven million 'page impressions' for papers such as *The Electronic Telegraph*, and four million for *The Times*. It also assesses actual user numbers for a minority of organisations, with ITN attracting the highest figure of these with 153,000 users per month (latest available figure, March 1998). Papers make their own claims too. *The Electronic Telegraph*[40] calculates that more than 60,000 people every day log on to its site from 200 countries around the world. The veracity of all such statistics, however, is not without question, for it is notoriously difficult to calculate electronic usage. Firstly, the definition of 'page impression' is misleading. A page can be made up of many images and more than one frame – and in standard transaction log software each is registered separately, artificially inflating usage. Secondly, as a study currently being undertaken at City University[41] is beginning to show, many 'users' – and the really big ones – are, in fact, web 'robots'.

## Conclusion

The survey of online newspapers outlined clearly shows the seriousness with which news providers are taking the new medium and the efforts being made to attract the online news reader. In some cases, such as the *New York Times'* educational supplement and the

'Euronet' initiative in *The Electronic Telegraph*, the amount of information included on the web version of the papers is impressive. If this practice is rather more widespread and comprehensive in the US, in other respects the US are even further ahead. Indeed, in some areas the UK appears to have eschewed the possibilities offered by the medium. Unlike the US, for example, where every major news event spawns a multitude of online bulletin boards, discussion groups etc., here in the UK the Net is still regarded as a passive medium, another outlet for news providers to tout their wares. Many papers, including *The Times* and *The Independent*, are even reluctant to exploit that most fundamental aspect of web technology, hyperlinking. Even internal page connections are lacking. Given the low take up of the Internet, and even lower online readership base, this reluctance may be understandable.

UK broadcasters, perhaps due to their familiarity with 'screen' media and with greater financial and (hence) staff resources, have adapted better, with the BBC facilitating discussion, incorporating audio and video material and providing external links. Whether online competition and/or the mass access to the Internet will galvanise print titles to embrace, however reluctantly, its true potential remains to be seen.

## References

1. Beckett, A. 'From press to click.' *Independent on Sunday*, Review section, 8 May 1994, p.106.

2. Sigel *Videotext: the coming revolution in home/office information retrieval.* (New York: Knowledge Industry Publications, 1980.)

3. Greenberg, S. and Lin C.A. *Patterns of teletext use in the UK.* (London: John Libbey, 1987.)

4. Teletext *About us http://www.teletext.co.uk/index/abou.htm* (visited 22 January 1999)

5. Bates, M. 'Editor's choice: newshound.' *Database* 20(2), 1996, p.96.

6. Nuttall, N. 'Thousands hail the paper that reflects our changing times.' *The Times*, Interface section 10 January 1996. Available at: *http://www.the-times.co.uk/cgi-bin/BackIssue?2632195* (visited 8 October 1998)

7. Times 'Internet Times at the top.' *The Times*, Features section, 13 August 1977. Available at: *http://www.the-times.co.uk/cgi-bin/BackIssue?2632195* (visited 8 October 1998)

8. MacArthur, B. 'I have seen the future, and it lurks in a laptop.' *The Times*, 31 July 1996. Available at: *http://www.the-times.co.uk/cgi-bin/BackIssue?2632195* (visited 10 October 1998)

9. Nicholas, D., Williams, P., Cole, P. and Martin, H. *The media and the Internet* (London: Aslib, 1998.)

10. Wall Street Journal. Site map *http://info.wsj.com/mediakit/tour/frames/sitemap_frames.htm* (visited 24 October 1998)

11. Sheppard, C. *An evaluation of British newspapers on the World Wide Web.* (London: City University, 1997.) (unpublished Msc dissertation)

12. Gill, C. 'Paper chase.' *The Independent*, Tabloid section, 14 October 1996, p.5.

13. Jamieson, T. 'Hot off the WEB.' *The Herald*, 22 September 1998, p.8.

14. Lloyd, C. 'Mission control keeps data flowing.' *The Sunday Times* Innovation section, 7 January 1996. Available at: *http://www.the-times.co.uk/cgi-bin/BackIssue?2632195* (visited 10 October 1998)

15. Preston, P. 'The whiff of JamJam tomorrow on the Net.' *Observer*, 11 January 1998, p.5.

16. Williams, P. and Nicholas, D. 'The Internet, a regional newspaper and an attempt to provide 'value added' information. *Aslib Proceedings* 50(9), 1997, pp.255-263.

17. Hanson, C. 'The dark side of online scoops.' *Columbia Journalism Review* May/June 1997. Available at: *http://www.cjr.org/year/97/3/scoops.asp* (visited 5 October 1998)

18. Nicholas, D., Frossling, I. and Buesing, P. '(Really) getting to grips with the Internet: what it has to offer in the way of newspapers.' *Vine* 104(29), 1997, pp.29-34.

19. Withey, R. (1999) Personal communication with the authors. [Richard Withey is Head of new Media at News International]

20. Katz, I. 'Online or not: newspapers suck.' *Wired* 9, 1977, p.50.

21. Hume, E. *Tabloids, talk radio and the future of journalism: technology's impact on journalism.* (Washington DC: The Anneberg Washington Program in communications policy studies of Northwestern University, 1995.)

22. Pavlik, J.V. 'The future of online journalism: a guide to who's doing what.' *Columbia Jour-*

*nalism Review* July/August 1997. Available at: *http://www.cjr.org/year/97/4/online.asp* (visited 23 September 1998)

23. MacArthur, B. 'Newspapers not yet a dying breed.' *The Times* Media section, 6 March 1998. Available at: *http://www.the-times.co.uk/cgi-bin/BackIssue?2632195* (visited 13 October 1998)

24. MacArthur, B. 'I have seen the future, and it lurks in a laptop.' *The Times*, 31 July 1996. Available at: *http://www.the-times.co.uk/cgi-bin/BackIssue?2632195* (visited 10 October 1998)

25. Erbach, G. *Subject: email correspondence to The Times*. Email to the author received 22 January 1999. [Gertrud Erbach is Reference Services Manager at News International]

26. Chapman, P. 'Developing This is the North East.' *Aslib Proceedings* 50(9), October, 1998, pp.264-266.

27. Nicholas, D. and Frossling, I. 'The information handler in the digital age.' *Managing Information* 3(7/8), July/August, 1996, pp.31-34.

28. Outing, S. 'A decade of online newspapers – lessons and trends.' *NetMedia 97 Conference Proceedings*, 1977. (unpaginated)

29. Rusbridger, A. 'Get the Net: whatever the opposite of an anorak is, I'm it. But in my simple, untutored way, I love the Internet.' *The Guardian*, 11 January 1999.

30. Hunt, R. Remarks made in a discussion forum at the annual conference of the Association of UK Media Librarians, 12 October 1997. [Robin Hunt is ex Head of New Media at *The Guardian* and now Director of Arhuss]

31. Eisenschitz, T. 'Internet law and information policy.' *Aslib Proceedings* 50(9), October, 1998, pp.267-273.

32. McAdams, M. (1995) *Inventing an online newspaper. http://www.sentex.net/~mmcadams/invent/* (visited 29 September 1997}

33. Griffiths, R. 'Video journalism – the lessons so far.' In *NetMedia 97 Conference Proceedings*. (London: City University, 1997.) (unpaginated)

34. Outing, S. 'The Net effect on journalism.' (Part 2) *E&P Interactive*, 23 February 1998. *http://www.mediainfo.com/ephome/news/newshtm/stop/st092398.htm*

35. Outing, S. (1998b) 'TV-newspapers combo beats individual small.' *E&P Interactive*, 23 February 1998 *http://www.mediainfo.com/ephome/news/newshtm/stop/stop.htm* (visited 20 January 1999)

36. Interactive Publishing Conference, Zurich (1998) Fifth Zurich, Switzerland Interactive Publishing Conference http://www.interactivepublishing.ch/conference.Info.interactivepublishing.iq (visited 20 January 1999)

37. ZDNET (undated) *World Wide Internet Usage Survey* http://www.zdnet.co.uk/zdm/netsurvey/page22.html (visited 19 January 1999)

38. Pew Research Center (1999) Change in News Consumption of US Users. *http://www.people-press.org/* (visited 24 January 1999)

39. ABC Electronic: *Electronic Newspaper Audits* http://www.abc.org.uk/electronic/ (visited 22 January 1999)

40. Electronic Telegraph ET shortlisted for a BAFTA *Electronic Telegraph* archive: 15.10.98 http://www.telegraph.co.uk:80/et?ac=000923575723312&rtmo=p1sSehpe&atmo=777777Ft&pg=/et/98/10/15/nbafta15.html (visited 20 January 1999)

41. Nicholas, D., Huntington, P., Williams, P. and Lievesley, N. (in press) 'Developing methods to determine the use of web sites: case study web newspapers.' [draft paper submitted for publication]

## Notes

[i] To be fair to *The Guardian*, however, it did alter its layout to carry news of the (original) Law Lords decision regarding General Pinochet's extradition, after the fieldwork for this paper was completed.

[ii] The graph produced by ZDNet shows, for example, a 5% decline in 'News weeklies', it is not made clear whether this signifies a decline in general readership of 5%, or a decline in Internet using readers, or, indeed, whether Internet users consult such magazines less, despite still buying them. Also, there is no explanation given as to how the decline is directly attributable to the Internet, as claimed. This is surprising, as the study was commisioned by such business heavyweights as Yahoo and Dell, for whom clear and accurate data would be a pre-requisite.

## Acknowledgements

The research on which this article was based was generously funded by Richard Withey, News International.

## Subscriptions

*Aslib Proceedings* is published ten times per annum and supplied to corporate members of Aslib. For 1999 the annual subscription rates are £170/$280 (£136/$224 to members of Aslib). For subscriptions and back issues please contact: Portland Press Ltd, Commerce Way, Whitehall Industrial Estate, Colchester CO2 8HP, United Kingdom. Tel: +44 (0) 1206 796 351. Fax: +44 (0) 1206 799 331.

---

Aslib, The Association for Information Management, is a world class corporate membership organisation with over 2000 members in some 70 countries. Aslib actively promotes best practice in the management of information resources. It lobbies on all aspects of the management of, and legislation concerning, information at local, national and international levels. Aslib provides consultancy and information services, professional development training, conferences, specialist recruitment, Internet products, and publishes primary and secondary journals, conference proceedings, directories and monographs.

Further information can be obtained from: Aslib, The Association for Information Management, Staple Hall, Stone House Court, London EC3A 7PB, UK. Tel: +44 (0) 171 903 0000. Fax: +44 (0) 171 903 0011. Email: aslib@aslib.co.uk *http://www.aslib.co.uk/*

---

Periodicals Postage Paid at Rahway NJ and at additional mailing offices.
**POSTMASTER:** Send address changes to PROCEEDINGS, c/o Mercury International, 365 Blair Road, Avenel NJ 07001

---

©1999 Aslib and contributors ISSN 0001 253X

ISSN 0001 253X
Volume 51
Number 5
May 1999

25/2/2000
wel

Aslib

# Proceedings

## New information perspectives

### Have you heard the library news? Personalised Net alerts for librarians

Robin Yeates

### Developing and testing methods to determine the use of web sites: case study newspapers

David Nicholas, Paul Huntington,
Peter Williams, Nat Lievesley,
Tom Dobrowolski and Richard Withey

### Information available on UK agricultural web sites

Gillian M. Craig

### On universality in productive and information technology

Julian Warner

Aslib THE ASSOCIATION FOR INFORMATION MANAGEMENT

*Aslib Proceedings* is a long-standing professional journal of international repute, which has a large and wide subscriber base. The journal brings currency, authority and accessibility to the reporting of current research, issues and debates in the broad area of information work. Above all the journal wishes to provide research and comment in a form that is easily and quickly understood: a fresh, rigorous, but unfussy, writing style is what is aimed for. The needs of busy practitioners (and academics) is very much in mind. Articles are now refereed to meet the growing need for accredited and authoritative information. To ensure that this does not result in an unacceptable loss of currency, articles are speedily refereed by the journal's large, eminent and multi-disciplinary Editorial Board (assembled for just such a purpose). In exceptional cases they may be referred to an external panel of consultants. The journal's monthly publishing pattern helps in maintaining the currency of reporting. In consequence the submission of research in progress articles is encouraged.

The boundaries of information work are rapidly changing and widening. *Aslib Proceedings* has a role to play in the repositioning of the information profession by encouraging submissions from information managers, providers, packagers and consumers of all kinds – and especially those from the related information fields of journalism, electronic publishing, communication and Internet studies. An interdisciplinary Editorial Board has been assembled to assist in the process.

Prospective authors should submit articles of 2,000-6,000 words in length. Articles should be accompanied by a 100-150 word informative abstract and the numerical form of referencing should be adopted. In style and tone, submissions should fall between that of the journal's sister publications: *Managing Information* and *Journal of Documentation*. Articles should be furnished by email or on disc to David Nicholas, Editor, Aslib Proceedings, Department of Information Science, City University, Northampton Square, London EC1V 0HB. Email: nicky@is.city.ac.uk. The Editor is happy to provide advice on an informal basis to potential authors.

**Editorial Board**
Dr David Nicholas, *Head of Department, Department of Information Science, City University;* John Akeroyd, *Director, Learning Resources, South Bank University;* Sarah Blair, *Head of Publications, Aslib;* Peter Chapman, *Information Manager, NewQuest North East;* Professor Peter Cole, *Head of Department, School of Journalism, University of Central Lancashire;* Dr Tom Dobrowolski, *Senior Lecturer, Institute of Information Science and the Book, University of Warsaw;* Harry East, *Senior Research Fellow, Department of Information Science, City University;* Mike Hyman, *Managing Director, Head Software International;* Robin Hunt, *Creative Director, AreHause;* Nat Lievesley, *Network Manager, Centre for Policy on Ageing;* Helen Martin, *Information Manager, Research and Information Unit, The Guardian*; Dr Ian Rowlands, *Lecturer, Department of Information Science, City University;* Richard Withey, *Director, New Media, News International.*

**Editorial Team**
*Editor:* David Nicholas
*Deputy Editor:* Ian Rowlands
*Editorial Manager:* Kay Nicholas
*Editorial Assistants:* Peter Williams (City University), Colette Coles (City University)

# CONTENTS

137-143 Have you heard the library news? Personalised Net alerts for librarians
Robin Yeates

144-154 Developing and testing methods to determine the use of web sites: case study newspapers
David Nicholas, Paul Huntington, Peter Williams, Nat Lievesley, Tom Dobrowolski and Richard Withey

155-166 Information available on UK agricultural web sites
Gillian M. Craig

167-171 On universality in productive and information technology
Julian Warner

# Have you heard the library news? Personalised Net alerts for librarians

*Robin Yeates*
*LITC, South Bank University, 103 Borough Rd, London SE1 0AA*
*Email: robin.yeates@sbu.ac.uk*

**Abstract** *The role of software agents in the provision of personalised news is explored in the context of the NewsAgent for Libraries project, which has been developed as a prototype and is about to launch a new free news alerting service for library and information staff. The issues that have arisen as the project moves from the research phase to service implementation are described, along with the prospects for the planned service.*

## Introduction

When most people talk of harvesting, if indeed they still do, it conjures up not merely images of routine hard labour, but also of close-knit communities, created by physical proximity and by the need to work together to survive. When some people now talk of harvesting, they refer to spiders and crawlers (not creepy-crawlies). They may even mention robots, harking back to a great novelty of the twentieth century. These people are software developers of the Internet, the new global explorers and shapers of landscapes. Most recently they have begun to witness the rise of human terms 'agents' and 'brokers' in place of robots. These agents, though, are far from human, although they act directly on behalf of human beings. They control other, subservient, pieces of software such as the robots.

This paper explores one function of the new software agents, their role in current awareness or alerting. In particular it considers what role software agents might play in helping librarians develop their careers, manage their libraries and network with others. Will it be possible in the future for agents on the Internet to foster virtual communities of librarians, to emulate the informal coffee-room gossip, the formal printed current awareness services and the invisible college, perhaps even to replace them?

## Virtual communities

Nobody ever sat down to design society. Librarians might well have been amongst the most resistant to the idea if they had done. In many ways, librarians were designed by society, or rather they have taken note of social behaviour and tried to add value to it through selectivity, acquisition, organisation and sharing. As the world 'goes digital' and the world of the written text begins to use the same communication channels as the broadcast media, librarians are beginning to play a part in shaping the digital landscape. In an era of rapid innovation, news and current awareness take on an extreme importance. We need to make intellectual connections and keep an eye on the overall trends and direction in which we are heading, without suffering from the information overload which can affect the less well equipped.

One current Internet trend is that towards 'Portals' or 'Virtual Communities', which go beyond the simple provision of material by a publisher to include sharing of electronic mail, discussion list comments, votes and preferences. A key difference from the already 'traditional' subject gateways and search engines such as Infoseek or directories such as Yahoo is the need for an individual to join. Although this usually involves nothing more onerous than choosing a username (any name will do in many cases) and a password, it actually requires the user to make what is clearly a very difficult decision. We know that at this stage most of those who visit the web site back away.

In most cases the user will receive little or no information that could not have been found in other ways, and the user may be required to 'visit' the service regularly to make it worthwhile trying to remember the password. Such an allocation of attention to a single service will only be warranted if it delivers the kind of familiarity and usefulness traditionally the preserve of one's circle of friends or everyday colleagues. Those who do join the 'club' will share much in common, therefore, including the need to be informed of the same information, and the opportunity to pass on information to others in the group confident that it is likely to be relevant in more cases than not. In theory some of this communication activity could be facilitated via the Internet, where geographical location becomes less important. Electronic mail, indeed, is usually the first Internet service to gain long-lasting popularity in a small group or organisation. The web is seen as a handy 'encyclopaedia' or passive reference work, but not widely yet as a collaborative or social medium.

## NewsAgent for libraries project

Librarians are inquisitive, acquisitive, and informative. How might they respond to the idea of an electronic current awareness service? A consortium of librarians and developers in the UK is seeking to find out, by creating an open standards based alerting service. It can be applied to the world of Internet resources and operated within the financial and practical constraints of all the stakeholders, including information providers, service operators and librarians as end-users.

NewsAgent for Libraries was originally funded as a two year collaborative project starting in April 1996 by JISC, the Joint Information Systems Committee of the UK higher education funding councils, under the Electronic Journals strand of the Electronic Libraries Programme (eLib). The details of its background and nature of the work completed during this phase have been described in a separate paper[1], and there is a project web site[2], but I include a brief description here, before discussing some of the issues that have arisen from the project.

Publishers in the field of library automation formed the core of the project team, including:

- LITC at South Bank University, which publishes *VINE*, and also *Library Technology* (jointly with the Library Association);
- UKOLN, the UK Office for Library and Information Networking at the University of Bath, which publishes *Ariadne*;
- University of Wales at Aberystwyth, Department of Information and Library Studies (UWA/DILS), from where staff edit the Aslib journal *Program* and the Bowker-Saur *Journal of Librarianship and Information Science*.

Each of these partners in the consortium also took on additional roles in the project, including project management and dissemination (LITC), metadata handling and pilot service operation (UKOLN) and user needs analysis, editorial issues and vocabulary control (UWA/DILS).

These partners were joined by CERLIM, the Centre for Research in Library and Information Management, originally at the University of Central Lancashire and moved in 1998 to Manchester Metropolitan University. CERLIM acted mainly as the internal evaluator of the project, setting and monitoring adherence to criteria in the light of the user needs analysis and its extensive evaluation experience.

Last, but by no means least, Fretwell-Downing Informatics (FDI) provided the main software development effort, although in the end UKOLN also played a significant part in creating, implementing and evaluating some aspects of the technology, such as the metadata harvesting system.

### *Stage 1*

Some nine months were spent setting up the project and gathering information from various categories of LIS personnel, including newly qualified librarians, subject librarians, researchers and academics, middle managers, senior managers and strategic managers of information services. The study identified the main problems faced by LIS staff in keeping up to date as lack of time, irrelevant and inappropriate information, information overload, not knowing what is available and where to find it, the volatility of information, and a lack of resources. More use was made of informal than formal current awareness services, and the idea of selective electronic alerting was apparently popular. During this phase we also commissioned a study on filtering agents by DJB Associates that highlighted both their potential and the possible intellectual property rights and commercial problems involved.

### *Stage 2*

During the second phase, from December 1996 to June 1997, we developed and tested an operational pilot service, using mainly a server at UKOLN to house personal subject profiles and registration information as well as the metadata harvester and associated data. Considerable effort went into designing a system which would be robust and which would fit well into a library systems environment. We were also keen to ensure that, if successful, the software could be marketed by FDI for other purposes beyond the NewsAgent for Libraries service. In this way we might have long-term support for the system and widespread acceptance of our metadata publishing and harvesting model, on which any such automated alerting system depends.

In retrospect, the NewsAgent architecture was probably too elaborate for our needs. We considered many long-term issues which could have been ignored had we simply wished to set up a single small-scale demonstrator. As a result, we constructed a complex set of software modules to manage and deliver the system, ultimately to send users regular, personalised, text-based, email alert messages and to offer a personalised web page. Some of these modules existed from other prior work, but the work of integrating them and configuring them for our purposes was very significant, and it proved difficult to maintain integration with some of the modules which were being developed for commercial marketing purposes at the same time.

### *Stage 3*

Within the time-scale of the initial two-year project, we were forced to severely constrain the number of iterative changes to the software design, so that we could complete a user evaluation of at least some aspects of the system. We tried to evaluate not merely the implementation, but also the whole concept of an electronic alerting service using email, the web and centrally stored personal profiles, which could be created and amended using an ordinary web browser. This was particularly difficult under research conditions. We were only able to muster a limited amount of content during the project, because we were uncertain of the nature and format of metadata we wished information providers to supply to us. Neither did we know the commercial conditions under which such a service might continue.

### Pilot system

During the initial two-year phase of the project, the service operated as follows. Anyone could join the service using an ordinary web browser by visiting the service web site and submitting a form with a user name and password of their choice.

The user then visited a Personal Profile web page to create a profile that would match with incoming news items. This was done by offering users a list of predefined Channels subdivided into Topics. Users could select one or more Topics by checking boxes. Users also selected, on the same form, the frequency for alerts to be generated. These could be sent via email Daily, Weekly or Monthly, with a further option Never applying when users only wanted to use the web for viewing alerts. A further option provided was to create one or more Personal Topics, which consisted of named stored queries. These Personal Topics, once created, were displayed along with the predefined Channels and Topics for selection or de-selection.

Once users had confirmed their profile, they would begin to receive email alerts from the next day and at the selected intervals thereafter. The alerts were in the form of text messages containing introductory and footnote materials referring to the purpose of the email and how to visit the project website for administration of the account, plus the alerts themselves. These consisted of a structured list of news item titles classified by Channel name. Each title was accompanied by a URL to the original full text, or at least an abstract in HTML of it. In this way users were alerted to events information maintained on the web by Thomas Parry Library, Aberystwyth, news items published on the web in HTML by LITC, and the body text of emails specially stored in HTML format inside the NewsAgent database, held using Oracle server software. Because we used the http:// prefix before URLs given in the email alerts, it was possible for a growing number of users to see this as a live link and simply click on it to retrieve the original texts. This very popular and practical feature relied on the capabilities of recent browsers and email clients to identify links within text in this way.

When users wanted to change their profile, in the light of the information they were receiving, they could revisit the NewsAgent web site to amend the frequency of delivery (thus receiving less or more alerts per email message), or the Topics or Personal Topics covered. Authentication for this was simply by requiring login using the previously selected username and password. Changes to the database, other than to these Profiles had to be carried out using a separate Windows client software package which used ODBC to communicate with the Oracle server, and this was therefore only installed at two user sites for the system administrators (in London and Bath).

## Alerting service issues

Evaluation of the initial phase of the project made it clear that further work was needed on several aspects of the service before it could be launched. These included some technical enhancements, mainly in the areas of query formulation and the user interface. More importantly we needed to examine possible business models which would allow continuation of the service, and operational arrangements which could support the preferred options. Finally, content has to be developed, because during the initial phase we had focused on achieving operational software and were not able to prove this sufficiently to proceed with marketing of the service to publishers and information providers.

### *Technical enhancements*

A small working group analysed the recommendations for improvements to the system and drew up a list of priorities. This was necessary because to achieve the full list of desirable features would have far exceeded our budget and probably would have made the service uneconomic. Our principal aims were to achieve an acceptable or attractive level of functionality by mid 1999 and a sound basis for future development if led by demand. We also sought to reduce the necessity for human involvement in the day-to-day operation of the system to an absolute minimum, to reduce operating costs. This means passing most of the responsibility for vocabulary control and effective configuration of profiles to the information providers and users respectively, although the system architecture means that we should be able to take advantage of progress in agent technology as it occurs.

A key decision taken was to manage the service from a single server instead of a fully distributed metadata content store, as had been the original intention. All harvested metadata would now be channelled into a single Oracle database, avoiding problems of registration, authentication and maintenance and greatly reducing costs and the development time-scale. Distributed Windows clients may still administer the server. This meant also that it was no longer necessary to use the Z39.50 information retrieval protocol to query the distributed servers. However, the development work is still being carried out in such a way that it should be practical to extend the NewsAgent software to a fully distributed environment. FDI have decided to develop the system with a view to marketing it independently of the NewsAgent for Libraries service, and may require this capability for some future customer.

Furthermore, it seems likely that we shall be able to develop metadata harvesting capabilities as standards and tools develop, so that data such as new book titles held in remote Z39.50 servers could be made accessible to the NewsAgent store. This means that content held on any web server with metadata in HTML, or other formats in future, can be integrated with NewsAgent without development of a network of fully distributed servers which might add to management costs and reduce the level of automation possible.

Two decisions were important from a software development perspective. One was a requirement from South Bank University to support an NT Server version. The other was FDI's need to productise the software so that it can be marketed and supported in the medium to longer term. Both these requirements have led to significant effort porting the system to a new software platform before the user enhancements are added. One benefit will be that an early demonstrator based on a portable NT version of the system can be created by March 1999, and by July 1999 there will be versions for both NT 4.0 Server and Sun Solaris 6.5.

From the user's point of view, the main improvement will be in the way profiles are set up. After experience of trying to use natural language queries as the basis of profiles, we now intend to offer field Boolean queries instead. The pilot system showed that email alerts are much more popular than personalised web pages, and more likely to be used regularly. However, users wish to keep email messages to a certain length and frequency, in which case there is no point in sending alerts by matching items with profiles through relevance ranking. We found that using natural language queries confused users and reduced relevance. Users expected their emails to contain ONLY items which were 'about' the concepts for which they searched, but many items included by the system algorithms were not perceived as relevant because there was no implemented cut-off point of relevance. For example, a search for 'library AND technology' in practice retrieved all items about 'library' OR 'technology' because once the Boolean ANDed terms had been applied, each term was also applied individually, albeit with the matching items being ranked lower. A field system will also allow users to select particular information feeds as well as Topics, although we have not yet decided to offer this.

### *Business model*

NewsAgent for Libraries was intended to be a self-supporting service, adding value to scholarly electronic publishing. During the trial phase, concepts such as 'push publishing' and 'information overload' have been invented, hyped, have waned and taken their place in the information landscape. In this landscape, users do not pay directly to be alerted to things. Often they do not pay for full text information itself, especially if it is perceived as the kind of material which would have appeared in magazines and newsletters or bought by libraries and used by readers for nothing. An alerting service appears to be much appreciated by users of web services as an added value feature. Much of this value stems from its ability to overcome the current network delays and cumbersome navigation options of many web services.

CERLIM has recently conducted an analysis of the possible business options for NewsAgent for Libraries as it moves towards a self-supporting service. A number of different models have been examined, including individual subscriptions, corporate subscriptions, piggy-back service, advertising, sponsorship and various kinds of public support. It is recognised that a mix of these basic approaches may be optimal. What seems clear is that the alerting function, whilst being technically complex to develop and to operate, is something that should be delivered as a part of a wider range of services to a group of users. These users will not pay specifically for the alerting service, but will receive it as part of a wider package of information. In this way, users of the existing service are regularly made aware of its benefits, but are also encouraged to stay with the virtual community, because they are still aware of external information resources that NewsAgent for Libraries can make available, yet need not leave the community to use them. In this respect, NewsAgent will provide a service analogous to those of the newspapers that highlight key headlines or leaders of their competitors. To test this model, we are therefore actively seeking support from organisations that wish to develop this kind of service to their members or subscribers.

### *Operations and support*

During the development phase of the project, UKOLN hosted the service in Bath, where they were able to work on its integration with other metadata handling projects, such as work on subject gateways. However, for the operational phase, it was deemed more appropriate for LITC to take over the service responsibility, since it has a role as a small publisher as well as an interest in cross-sectoral LIS activity.

Using a portable NT system, we intend to demonstrate the system to potential information providers and sponsors from March 1999 onwards, with a view to launching a service based on a Sun Solaris server at South Bank University from mid summer 1999. It will be supported by LITC with marketing assistance from CERLIM, some technical support from FDI and metadata handling support from UKOLN.

### *Content*

Lastly, but for users the most crucial issue, is the content. In the case of NewsAgent for Libraries it is recognised that decisions on content need to be taken with a strong editorial lead, as if we were publishing a news magazine, even though the NewsAgent consortium does not expect to own much of the full text content. It will be important to take a dynamic view of the required content and to select those sources expected to be of most interest to registered users. These users will have joined not as a result of knowing the value of the content, but as a result of our marketing activities or out of curiosity. We know from our research that we must provide a reasonable coverage of chosen information sectors of interest to our audience. We also know that disappointment will be felt if the user receives no alerts, because a profile (stored query) expresses a desire for information just as much as a desire to filter other information out.

The whole area of user profiles is ripe for further study and development, because it requires an understanding of social behaviour, language, communication and feedback systems. We have yet to determine the level of duplication of source information that is acceptable (so that users may compare different news sources for the same news, for example), and also how best to maintain high quality. Experience from searching, or browsing search engines, suggests that including all sources and then providing filtering capability can provide a useful 'last resort' finding service. Can the same principle be applied usefully to alerting, where the output flow needs to be reasonably consistent, constant, quite restricted in quantity, yet relevant?

For practical purposes we shall be concentrating to begin with on assembling a high quality collection of material from experienced publishers sprinkled with some new sources perhaps. Information providers in many cases should not have to pay to alert the LIS community to their content. They will need to adopt appropriate metadata publishing procedures in collaboration with LITC, whether specially for the project, as may be the case for small publishers, or as a part of their normal electronic publishing, as is more likely for larger publishers. These publishers will gain an outlet channel for their content, in return for which users will receive improved access to information on a timely basis.

Benefits should come primarily as a result of the collaborative cross-sectoral use of metadata that the system promotes, but also partly as a result of the open, managed integration that it should also achieve. We shall deliver a press alerting service with improved links to original sources as well as users, an SDI service where originators can monitor the effectiveness of their own indexing efforts and a marketing channel where users are in command.

### Conclusions

> 'The introduction of these news and alerting services into a library environment will have the positive effect of freeing up staff who are continually pressured with requests for current news. But there will always be a need for selective dissemination by human subject experts.'[13]

NewsAgent for Libraries is seeking on the one hand to harness computers to process increasing amounts of text and present an organised view of it to users at appropriate intervals and via appropriate telecommunications channels. On the other hand it hopes that by eliminating new human intermediaries who might be needed by libraries to provide current awareness in cyberspace, we can improve the effectiveness of human beings, whether as authors, publishers, indexers or readers.

Some worry about the 'dumbing down' of news services through 'personalization' such as this. Robin Hunt, for example, has thoughtfully analysed the problem in terms of 'its pros and cons, its hermetic, blinkering vision; or its liberating self-centred *usefulness'*. Of portals, he says 'speed and change is on their side as they march onwards encompassing and appropriating news. These organisations, not traditional publishers in any sense, accept news as part of a much broader package – called living'[4].

So, we are back in the real world, if not the rural idyll but one where electronic combined harvesters might feed our minds. We shall perhaps be able to facilitate informal gossip, if that is what users want. After all, journalists may well be feeding the service alongside colleagues and webmasters, academics, marketers and product suppliers. But we shall certainly be able to offer a new channel for LIS dissemination, and one that will place librarians firmly in the vanguard of current awareness, where they will naturally want to be. Their replacement in the wider world, if it ever happens, should then come as news to them before anyone else.

## References

1. Tedd, L.A. and Yeates, R. 'A personalised current awareness service for library and information services staff: an overview of the NewsAgent for Libraries project.' *Program* 32(4), October 1998, pp. 373-390.

2. NewsAgent for Libraries web site *http://agent.sbu.ac.uk/newsagent/*

3. *News and alerting services on the Internet*, Charles, S. LITC Report No.11. (London: LITC, 1998.)

4. Hunt, R. 'Dumbing down: the only way is up.' *Aslib Proceedings* 51(2), February 1999, pp. 46-53.

# Developing and testing methods to determine the use of web sites: case study newspapers

*David Nicholas, Paul Huntington, Peter Williams, Nat Lievesley, Tom Dobrowolski and Richard Withey*

*Internet Studies Research Group, Department of Information Science, City University,
Northampton Square, London EC1V 0HB
Email: nicky@soi.city.ac.uk*

**Abstract** *There is a general dearth of trustworthy information on who is using the web and how they use it. Such information is of vital concern to web managers and their advertisers yet the systems for delivering such data, where in place, generally cannot supply accurate enough data. Nor have web managers the expertise or time to evaluate the enormous amounts of information that are generated by web sites. The article, based on the experience of evaluating The Times web server access logs, describes the methodological problems that lie at the heart of web log analysis, evaluates a range of use measures (visits, page impressions, hits) and provides some advice on what analyses are worth conducting.*

## Introduction

Users (and uses) are being drawn to the web like bees to honey, and this is happening globally. With information retrieval (formal and informal), communication and business being increasingly conducted via the web – and with everything (in theory, anyway) open to routine monitoring, we are in an excellent position to examine information behaviour that has never before been put under the microscope – or certainly never so comprehensively or on such a grand scale. There is data on web newspaper 'readership', and it is an improvement on what is available on hard copy newspaper usage, but it tends to be very rough and ready[1]. Yet strategic planners, marketing departments and advertisers in the media are demanding accurate and detailed information. Where systems for collecting data exist it is all too often a case of:

a) the software not being up to the job because of the sheer size and complexity of the files (leaving a trail of misleading information in its wake); and/or,
b) few people having the time, knowledge or motivation to interpret the data (most web managers are too busy developing new projects to take the time to consider how the existing product has gone down).

The danger in all this is that large sums of money are being invested and strategic plans developed, at best, on the basis of poor and inadequate information, and at worse, on anecdote, wishful thinking and on past precedent. The companies that recognise this are turning to independent auditors to obtain help[2].

Ironically, this situation is actually relished by some – the so-called 'creative thinkers' of the media world – who see their cyberfantasies being punctured by the reality of the data or, more likely, fear the serious political fallout that would occur as a result of its release. This is best illustrated by an interview researchers conducted with the new media director of a major UK broadsheet a few weeks before it was to announce a multimillion pound investment in its web site. After explaining to the new media director what would be forthcoming from a log analysis of his site's logs, he wryly smiled and said that he would not want to know such detail; having such accurate data would cramp his editorial style and it would create all kinds of personal problems for him. For instance, what if the data showed that people never read the editorials but only sport? Such information might upset the journalistic apple cart, triggering off departmental in-fights.

However, without detailed and comprehensive use data many strategic questions cannot be answered until, perhaps, too late. Such questions as: will news migrate from hard copy to the web and what does that mean for the hard copy newspaper as we know it, and will the web attract a more youthful audience to the news?

## Research context

Within the information profession there is a good deal of experience in analysing what are termed 'transaction' logs, albeit those of OPACs and commercial online services[3,4,5]. Much of this knowledge appears to have failed to reach the web world and inform deliberations there – and information scientists themselves have been slow to play their part in web log analysis[6]. While web logs are different in their structure and detail – and come with their very own set of problems – there are many similarities with OPAC logs, for instance; and, after all, they are all trying to record the same thing – use. The work that has been done comes mainly from the United States. *Clark* was developed by David Clark[7] as a 'public domain, customisable, multi-platform, research oriented program' for the very purpose of collecting data in order to assess the effectiveness of WWW documents by creating transaction logs from HTTP server information.

A similar idea was behind the Archimedes Project[8]. Archimedes provided information about the Engineering Library at the University of Michigan. It also monitored users and generated transaction logs and various statistics. In 1994 the system was converted for use on the web, and a year later Archimedes terminals were replaced by a dedicated Netscape station set to the library's new WWW home page and information system. As Carter points out, 'it was the desire to be able to gather the same transaction information and statistics as was possible with Archimedes in the new WWW system which was one of the main impetuses for this project'.

More recently the University of Connecticut has reported on its 'MAGIC' (Map and Geographic Information Centre) transaction logs, set up to measure access and use of web accessible geodata. Log files were examined for four years by the project team, who determined that information could be collected about the following areas: user identity and location, files requested and navigational behaviour. McGlamery[9] claims that enough information was provided to 'know more about the virtual users than real-world ones'. Nevertheless, he acknowledges that face-to-face and telephone contact are also necessary to build a profile of the MAGIC user community. A point echoed by Kurth[10], who explains that although 'Transaction log data effectively describe what searches patrons enter and when they enter them, but they don't reflect, except through inference, who enters the searches, why they enter them, and how satisfied they are with their results.' Also, errors in log analysis can arise through limitations of the online system and the inability to isolate and characterise individual users.

Using the methods and techniques successfully developed to understand users/use of 'oldline' hosts and databases, and adapted to operate in the web environment, City University's Internet Studies Research Group are evaluating the use/users of *The Times/Sunday Times* web site. Working on enormous data sets the Group are discovering: how much the site is being (genuinely) used, how much use is growing by, who the main users are, where they come from, what do they do when online, where they visit (sport etc.), how long they spend there and much more. Some of the data emerging from the study appears to run counter to perceived industry wisdom. It is quite possible that some people are misreading the situation – especially in regard to the volume of use, and will have their fingers very badly burnt. Using the experiences of *The Times* study this article sets out the methodological problems that confront those wishing to analyse and evaluate web log data.

## Consumption/activity measures

Bereft of revenue figures (very few media sites charge) everybody involved with the web (web managers, newspaper proprietors and their shareholders, advertising agencies, sponsors etc.), are looking for measures of consumption, interest or activity, if not success or satisfaction. Some, rather crude, measures are being imposed on the industry by audit bureau, who are attempting to produce a common credible standard of website measurement, with a view to driving up advertising revenues[11]. Essentially consumption, of information in the Media's case, is measured in two ways:

1. by the number of *users* who avail themselves of the service; and,
2. by the amount of *use* made of the service.

There is a need to employ both measures, because employed on their own some big questions go unanswered. Thus you may have many people signing up to a site but not using it very much, lots of use but the responsibility of just a few 'users' (search engines, robots), lots of people making light, accidental and incidental use of the service (surely a characteristic of the web with its false trails, dead ends, thousands of hits). Use and users are then complimentary activity measures and should be used in conjunction, although we shall argue later – and the industry has conceded as much – that use plainly forms the bottom line in all activity calculations.

### *Use*

Use is the measure adopted by most web site managers, and auditors when they come to assessing (and publicising) the popularity and success of their sites. In fact two measures of use are commonly employed. The first is the 'visit' – calculating the number of visits made to a site, rather than the number of visitors. But what precisely constitutes a visit? Visits are in fact an old friend, the search session, which is denoted by someone or something arriving at the site, searching and then departing the site; ironically, as we shall learn, something that is hard to determine in respect to the web. Of course, a visit might entail nothing more than a stop on the way to somewhere else, or worse, an unwanted visit (as a result of an unhelpful link). Hence the second, rather more precise measure, page impressions downloaded or pages viewed – a measure sometimes wrongly referred to as 'hits' (see below). Page impressions are used to describe how active the user is during their visit. The difference, and relationship, between these two measures is best illustrated by some figures IBM released regarding the 'use' of the site set up for Garry Kasparov's chess confrontation with their Deep Blue computer. According to IBM they received 3.5 million visits, but these visits resulted in 65 million page impressions[12]. Page impressions is in fact the metric adopted by ABC, the newspaper readership audit company, to monitor media web sites. Of course, hard copy newspaper use is generally measured by another criteria, that of individual sales. A fairly hard and unambiguous measure you might think, but not so, for there are promotional and discounted sales (to airlines for instance) and, of course, a household/library will typically buy one newspaper but more than one person will generally read it (data newspapers collect as well). Web statistics, for all their faults, potentially offer more accurate and comprehensive consumption data than the industry is used to.

Page impressions might appear to be a fairly concrete and problem-free use measure, but they are not. Problems arise from three factors: the aforementioned inaccuracy of web searching – 'to assume that 'hits' [pages viewed] reflect user interest is to assume that users are as interested in not finding the information they need as they are in finding what they are looking for'[13]; the fact that to reach a desired page you might 'view' many intervening pages of various degrees of relevance (irrelevance); and the way a viewed page is recorded in the logs. Thus, in regard to the latter, a page downloaded from a web site may be represented in a log file by a single line or many lines depending on the number of frames/graphics represented on that page, with each frame/graphic giving rise to a line (i.e. there is not a one-to-one relationship between a page impression and a log line). These problems are discussed in detail later but for the metric to work, page impression calculations should exclude graphics, banners and menu frames, invalid bookmarks and other forms of noise. Lines on their own hardly constitute even a very rough and ready measure of use. Thus the analysis of *The Times'* web logs shows that just 17% of the lines recorded in the logs are valid. It is generally line counts that are being referred to when the word 'hits' (as vague a term as you are likely to encounter in the minefield of web evaluation) is used to describe use of a site – each line being a hit. As one commentator put it: 'counting hits is a meaningless currency'[13]. There is, however, another consideration: what of 'pushed' pages or client 'pulled' pages? As far as ABC are concerned they do not count. A valid page impression is defined as that which results from the user making an active request – it has to be triggered by a human. Further, any page that is automatically refreshed should only be counted once.

There is another, not so commonly used, measure and that is the amount of time spent online. A measure (once) much favoured by the commercial online (oldline) industry. The length of time a user spends on a visit or examining a page, for instance, is plainly of interest to a sponsor or advertiser – a useful way, perhaps, of distinguishing between surfers and searchers, intended and accidental users. The measure is somewhat devalued by the fact that time calculations contain two essentially very different components: processing or downloading time (download speeds differ from computer to computer, and vary according to the complexity of the page and the time of day), and searching time – and you cannot distinguish between the two in the logs. In some cases processing time might constitute a tiny proportion of the overall time, but in other cases it might constitute the majority. Also users, largely free from the disciplines imposed by connect hour charging and telephone charges, may idle away their online time and remain connected to the site even when they have finished searching. Nevertheless the measure has a growing body of adherents[14] no doubt dispirited with the other imperfect measures.

This leads us to the problem of defining the duration of the visit. The time the user logs on to the site is recorded. And each time the user navigates to a new page this is recorded in the log file. The lapse of time from the downloading of one page to the next defines the length of time spent on that page. However, as we have already explained, that time is not strictly time spent viewing the page and its contents, because there is no way of knowing whether, for instance, the user went for coffee during this time. This problem in part can be overcome. A site is composed of a number of pages, and pages at a site usually conform to a site style length. With a reasonably large population of users (virtually guaranteed with the web) we can calculate the average time spent reading a particular type of page, be it a new story, obituary, leader, and so on. Given this we can get a better idea of those users who have: a) clicked through without reading, b) those who have read the article, and c) those who have gone for a coffee.

Determining when a user leaves (exits) a site is problematical as access logs do not record when the user leaves, there is no sign- or log-off message. We are left merely with a 'long' break after the 'last' page view, and, because of this, there is no way of knowing how long someone spent reading the last page. It can be difficult to distinguish between the end of a session and the resumption of an existing session after a long break. Indeed, it can even be difficult to spot a user's arrival. This arises because of the 24 hour, uninterrupted and global, nature of the logs. Thus, for instance, if one particular day's visits were being monitored you would need to examine the previous days log's to see whether any searches were a hang-over from the previous day. Fortunately an exit (and arrival) may be determined by other means as well, in particular, how the user behaves when in the site. Thus many users arrive at the home page. In the case of news-related sites there is a very good reason for doing this, for it is the home page that will hold the current-breaking new stories. This is the page that will be bookmarked. Clearly some users bookmark other pages but then this characteristic identifies a particular type of user. Similarly there is compelling evidence to indicate that users like to leave after visiting a particular page. The data gathered from *The Times* web site shows that the obituaries page is one such page, sport pages another. Thus, although theoretically the difference between a long coffee break and an exit may seem difficult to define, statistically this difference can be modelled.

### *Users*

All kinds of methods have been used to obtain web user (visitor, audience) figures. Of the concrete measures the Internet protocol (IP) addresses left by the user in the web logs and subscriber lists/databases, where they are a condition of using the service, have both been turned to. However, neither of them are without reproach and it is generally true that accurate user data is hard to come by[3]. The biggest problem in both cases is resolving use down to the individual. Take IP addresses first. It is difficult to determine whether an IP address represents a computer being used by a single individual or a mainframe supercomputer used by many individuals, or indeed whether the individual is a 'person' or robot. There are also the problems that arise from the fact that when a dial-up account is granted a floating IP address is given, and when users come through firewalls they all appear to have the same address: the address of the firewall system.

> 'The consequence of these problems is that when someone visits your site two days running they might have a different IP address each time. Or two people visiting at the same time might share the same IP address.'[15]

There is another major problem associated with IP addresses – they are not easy to translate into (domain) names of users – something we shall return to later.

Subscriber or membership lists on their own will only disclose that someone is a potential user. Subscriber numbers are trumpeted by web owners but the reality, if known that is, is that often a tiny percentage of subscribers actually use the service. However, used in conjunction with use data, subscriber lists can be valuable: clothing the analysis with many personal attributes, like the gender, age, occupation and location of the user. But for this to happen there has to be some method of linking the two sets of data together – use and user data – and that link is often not there. *The Times* web site had just such a database (containing 1.5 million registered users). Its value, however, is limited by the fact that there is only a weak link between it and the log files. A link can only be established by relating the IP address with the email address given by the subscriber. Subscriber databases also have to be cleaned before they can be used as users appear to delight in providing false information – there are a large number of Tony Blairs and Mickey Mouses who use *The Times* site for instance. And there are quite a few 'duplicate' entries – people registering more than once because they forgot their password. It can still be a problem linking use to an individual. Thus even in the case of a named home user it must be the case that their login name and password will be used by more than one member of the household – in the case of an adult their partner or teenage children. And there is no way of knowing.

Of the 'softer' methods telephone and face-to-face interviews, mailed and web questionnaires are sometimes used but are generally insufficiently targeted to yield anything useful (on their own). As an adjunct to logs these methods can provide valuable background or causal data (e.g. information absent from the subscriber database or access logs, reasons why a particular search was conducted).

### *Over/under reporting and the phenomena of caching*

We have already seen how inaccurate individual use/users measures can be. However, there are also some general factors at work that distort all measures. There is one factor (caching) that leads to under-reporting and three (robots and spiders, internal traffic, and redirections) that lead to over reporting of the data. In all but the case of caching these factors are measurable and can be allowed for. Unfortunately the impact on the figures of caching is probably the greatest of all.

Caching is the process whereby the content of a web site is downloaded from a server once and then stored, so it can be served up to many users a multiple number of times. Caching results in usage statistics being underreported as it makes available proprietary pages to third parties. Technically – in the UK caching is a breach of copyright law and can be likened to wholesale and unbridled photocopying, but it is something that goes on all the time and is difficult to police. The situation is complicated by the fact that it is legal in the United States, where caching is considered a solution to the growth of network traffic (and the slowing down of response times that results). Caching protocols do exist. These basically state that cachers should provide full lists to the original site owner. However, there is little inducement for information 'providers' (in its widest term) to follow protocols – after all, should they do so, they would be handing out commercially sensitive data.

### *Access logs*

The usual way to report on visitors and usage patterns on World Wide Web sites is to examine the access logs representing records of files downloaded from the host computer to the user's computer. Access logs consist of a very large number of lines detailing the page impressions downloaded. Each line representing an access or 'hit' in some people's parlance. Typically the content of these lines will give limited identification of the user, the time and date of the download, the path and files downloaded, and number of bytes downloaded. As mentioned previously it is a mistake, although one repeatedly made, to consider each line of a log file as a page impression because a page can be constructed with graphic elements and a number, if frames are used, of individual files. To 'read' a page the user could have downloaded any number of graphic and composite files. Hence a page impression might be recorded in a log file by anything between 1 to 50 downloaded lines. This of course can lead to a huge inflation in the 'hits' count if a simple line count is made – something it has to be said that newspaper proprietors, with one eye on the advertiser/sponsor are not too unhappy about.

Unfortunately for the analyst the lines in an access log cannot be automatically transferred into either a database or a spreadsheet application. The problem this causes can be best understood when one realises, for instance, that *The Times* web site log generates up to 200 million lines a day. There are essentially two ways of analysing log files. Firstly, they can be analysed with the aid of 'off-the-shelf' software, such as WebTrends or Analog. Secondly, the analysis can be conducted semi-manually, parsing log files from their relatively unfriendly state into something that can be automatically processed by standard statistical packages or databases. The aim in both cases is to produce a useful descriptive analysis and a whole range of monitoring measures. However, 'off-the-shelf' software analysis tools do have serious shortcomings, the program decides on the information reported and handles what constitutes a 'visit' quite differently: some will include graphic downloads and most cannot distinguish between multiple downloaded lines due to the presence of pages with frames. The advantage of parsing log files is that it provides much flexibility, though the lack of a GUI interface makes the procedure unfriendly. A further disadvantage with 'off-the-shelf' programs is that they can rarely handle very large log files – and, newspaper log files are typically very very large.

### *Processing of web logs*

Typically, web log analysis consists of a number of steps:

1. the collection of log data which is automatically recorded by the server;
2. replacing or deleting any characters that might cause problems as the data is moved from one computer platform to another, i.e. from a UNIX environment to a Windows environment;
3. the stripping out of lines with graphic images and additional lines that result from frames – in the case of *The Times* logs these amounted to 83% of all lines;
4. if the files are very large then a random sample of lines will need to be selected, or an alternative selection procedure can be implemented – such as picking out all lines related to a particular user or to particular page;
5. the parse conversion of the remaining lines into a spreadsheet or a database useable format;
6. the analysis and further parsing and manipulation if necessary.

Parsing is a sequence of operations that replace, add, change or delete characters in a file. A trivial example of a parse function is the 'find and replace' function available on most word processing programs. Many programs have extensive parse routines available to the user as commands. However, it is usual to design a purpose built program. The code of the program manages each line of input and makes the requested changes and then writes the line to a new file. Parsing allows log files to be examined line by line and for changes to be made. It is fairly easy at this level to adapt the program, hence the process becomes flexible and tailor-made for the purpose.

### *The contents of the access log file*

There are a number of different formats to log files. The most common log format is probably the NCSA Common Log Files format. However, the demand for log information has resulted in the development of extended log formats: the W3C Extended Log File format and the Microsoft IIS Log File format are two such developments.

A typical line of a log file will look something like the following:

```
163.1.172.5 - - [11/Apr/1998:00:12:27 +0100] “GET
/standing/tim/worldback2.gif HTTP/1.1” 200 71
```

The digits *163.1.172.5* record the clients Internet Protocol (IP) number. By itself the number is not particularly useful, although the number can be looked up (by the parse program) from reverse Domain Name Server (DNS) look-up information made available by InterNIC.

While IP addresses can be resolved into their domain name equivalents by using 'Look-up' software, which uses databases of registered users to match the IP number with the domain name, and then calculates their longitude and latitude, there are problems associated with the whole process. Firstly, the location is only accurate to city, rather than organisation level and even this is only available in the case of US servers. Canadian sites are only resolved to their province, and other non-US sites to the country's capital. All UK servers, therefore, are listed as being located in London which is obviously less than helpful. Secondly, the location of the server is itself not necessarily an accurate reflection of the physical location of the users accessing the Internet through that gateway. Indeed, we have encountered traffic being routed through Japan that are thought to be from users in the United Kingdom. It is probable that personal users such as those with their own ISP accounts who use the Internet from home and who are therefore remote from a server, are (unbeknown to them) connected to the Internet some considerable distance from where they are actually situated. University users, on the other hand, enter the Internet from an internal gateway connected to a university LAN. This is also the way data traverses the Internet. Perhaps ironically, it is the sophistication of the technology that clouds, rather than reveals, the data here. With the Internet dependent on a packet switching system, traffic can not only be directed along the least congested routes, but also split up and routed down various pathways, to be reassembled at the destination.

Finally, the domain name itself has to be resolved one step further into a meaningful site for any real data to be accrued from it: 'inet2.tek.com', for example, is not particularly information rich. It should be possible to automatically generate data from such domain names. If not each address has had to be manually entered into the 'Go to' data entry box on a web browser and then accessed (or 'visited'). Even after much time has been expended on entering the data and waiting for the connection, there will still be prefixes to the addresses which cannot be resolved – perhaps as many as 30%. Thus, to take the same example, 'inet2.tek.com', may be '*http://www.inet2.tek.com*', '*http://www2.inet2.tek.com*', have another prefix, or simply be '*http://inet2.tek.com*'. In a URL, only the 'http://' is standard, not, as is sometimes thought, 'www'.

The segment *11/Apr/1998:00:12:27 +0100* provides the date and time of use and the time zone of the user compared to GMT. Time and data analysis rest on this data. Download time is of interest – how long a web page takes to load. Users access sites with different bandwidth constraints (e.g., 9600, 14.400, 28.800 bps, ISDN, Ethernet). A page that takes six seconds to download for one user may take six minutes for another. *GET* describes the method by which the line was downloaded onto the clients computer – methods include: GET, HEAD, POST, PUT, or DELETE.

The part of the line *standing/tim/worldback2.gif* records the name of the file downloaded, in this case a gif image, and the full path name from the document root. This is where the 'subject' information resides and it is the final part of the line that is significant – the destination file. *HTTP/1.1 200* – is the server status. *71* – refers to the size of the file downloaded to the clients machine, measured in bytes.

Other characters in the lines of the log file, such as square brackets, dashes, semi and full colons and backward or forward slashes are of no value and are in fact a hindrance when loading text lines into a database or a spreadsheet. Programs can operate on a variety of delimited text – the most usual being comma delimited. Here each field of information is separated by a comma. It is the parse program that can delete unwanted characters and insert commas at the beginning or end of each desired field. The parsed and processed file (see Table 1) can then be imported into a program.

There is more to log analysis than deciphering the line. Most importantly the 'threading' of the users have to be undertaken – plotting the path (thread) through which a user navigates a site. In this regard the following need to be considered.

*Entrance*

Where the user enters the site. The percentage of users who, say, enter from the Home page, as opposed to directly accessing a page within it. This can help determine where to put an announcement (e.g. regarding new content) or an advertisement.

*Exit*

The page(s) from which users most frequently leave a site. It could be that a large number of users leave a site on a specific page. This alerts a web administrator to re-examine that page. Maybe a link is down, maybe a graphic is taking too long to load, maybe the content needs some work.

The software for statistical analysis of *The Times* log files was designed using custom PERL and CGI scripts.

*Table 1. The parsed and processed file: a sample.*

| *IP address* | *date of of search* | *hrs.* | *mins.* | *secs.* | *GMT offset* | *destination page* | *HTTP version* | *Service status code* | *Bytes sent* |
|---|---|---|---|---|---|---|---|---|---|
| 134.117.1.22 | 11/Mar/1998 | 00 | 20 | 49 | 0000 | regis | 1.0" | 200 | 926 |
| 209.63.114.28 | 11/Mar/1998 | 00 | 24 | 31 | 0000 | home | 1.1" | 200 | 2801 |
| 195.44.0.224 | 11/Mar/1998 | 00 | 26 | 11 | 0000 | front | 1.0" | 200 | 1424 |
| 160.96.179.5 | 11/Mar/1998 | 02 | 20 | 07 | 0000 | fpcon | 1.0" | 200 | 8880 |
| 153.35.110.150 | 11/Mar/1998 | 02 | 25 | 45 | 0000 | timbi | 1.1" | 200 | 9531 |
| 207.102.33.157 | 11/Mar/1998 | 02 | 26 | 16 | 0000 | timnw | 1.0" | 200 | 9930 |
| 166.72.168.140 | 11/Mar/1998 | 02 | 41 | 34 | 0000 | timfg | 1.0" | 200 | 9151 |
| 203.10.130.1 | 11/Mar/1998 | 03 | 21 | 33 | 0000 | timnw | 1.0" | 200 | 14442 |
| 198.7.150.54 | 11/Mar/1998 | 03 | 24 | 49 | 0000 | home | 1.0" | 200 | 2801 |
| 198.53.4.182 | 11/Mar/1998 | 03 | 30 | 04 | 0000 | timin | 1.0" | 200 | 13581 |
| 202.242.209.55 | 11/Mar/1998 | 03 | 30 | 59 | 0000 | timin | 1.0" | 200 | 10896 |
| 152.163.197.39 | 11/Mar/1998 | 03 | 34 | 33 | 0000 | front | 1.0" | 200 | 1424 |
| 194.168.54.21 | 11/Mar/1998 | 03 | 44 | 17 | 0000 | timin | 1.0" | 200 | 14921 |
| 203.10.76.34 | 11/Mar/1998 | 03 | 53 | 57 | 0000 | timop | 1.0" | 200 | 7262 |
| 131.236.1.5 | 11/Mar/1998 | 04 | 31 | 15 | 0000 | timfg | 1.0" | 200 | 11366 |
| 202.80.18.109 | 11/Mar/1998 | 04 | 38 | 28 | 0000 | puzzl | 1.0" | 200 | 2902 |
| 205.238.244.83 | 11/Mar/1998 | 04 | 38 | 35 | 0000 | timwe | 1.0" | 200 | 9367 |
| 207.194.73.105 | 11/Mar/1998 | 04 | 41 | 58 | 0000 | timsp | 1.0" | 200 | 8484 |
| 209.51.93.95 | 11/Mar/1998 | 04 | 54 | 15 | 0000 | timfg | 1.0" | 200 | 9153 |

### Types of analysis

There are a whole variety of analyses that can be produced from web logs, some of which have already been referred to. What follows is a comprehensive description of all the analyses.

#### *Simple use measures*

The analyses that suggest themselves are:

- how many *users* (individuals or corporate) use the site per hour, day, week, year;
- how many *visits* users make per hour, day, week etc;
- how many *pages* uses view per hour, day, week etc. (See Figure 1 below for an example);
- how many *hours/minutes* the site/page is used per hour, day, week etc.

As well as providing general activity data, such analyses can provide: a ranking of users, the peaks and troughs in use, overall trends in use (increasing, decreasing), drop off rates for users, and most importantly of all, an idea of how dynamic the data is (i.e. how fast changing it is).

#### *User characteristics*

The type and the nationality of the organisation to which the user belongs are the main characteristics that can be gleaned from the logs. However, as we have said earlier, nationality can be difficult to determine, with users signing up to service providers in foreign countries to circumvent legal problems or reduce costs. The fact that the web has a truly global audience, and that most hard copy newspapers certainly have not, means that nationality data are of special interest.

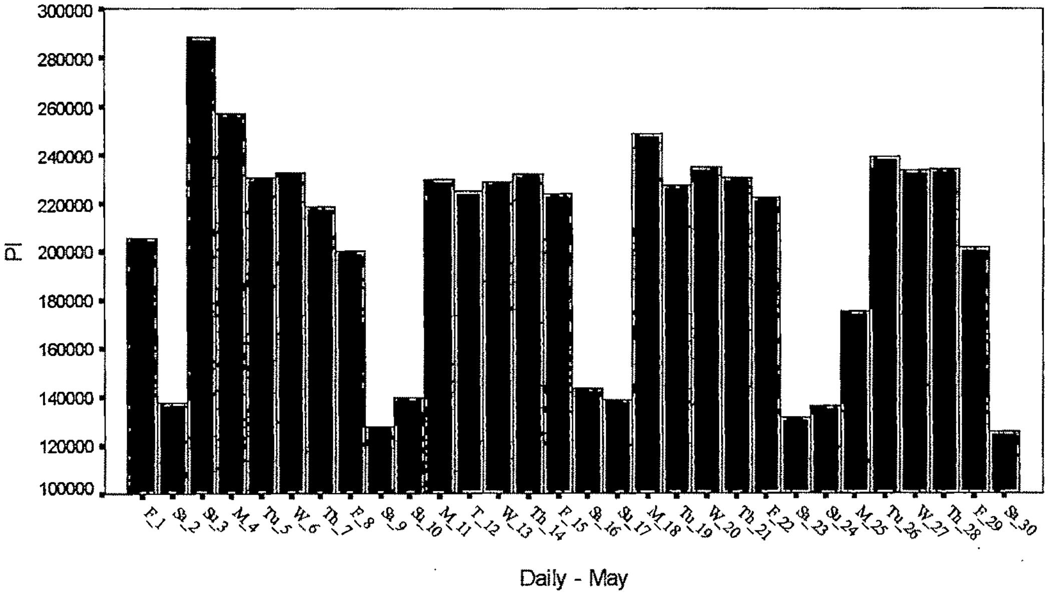
From the subscribers database comes the real demographic detail. Thus *The Times* database contains information on gender, age, and geographical location (postcode).

#### *Searching characteristics*

*Subject*

What subjects people search have always fascinated information scientists. As users of *The Times* site cannot input keywords of their own when searching the site (there is no search engine – instead they click on a menu or icons), it is not possible to determine exactly what they were looking for from the logs. Furthermore

*Figure 1. Daily use The Times, May*



their searches are constrained by what is on offer and although newspapers are encyclopaedic in content, the web content pages/maps do not reflect this. It is a pre-coordinated type of search and the success in finding what you want depends heavily upon the site designer having allowed for your approach. What information we have then is the pages that they go to (Figure 2). This of course provides us with loose and general subject information. In fact not quite so loose, for a typical site like that of *The Times* contains about 500 pages. However, many of these pages, being single story pages, are affiliated with other pages, so the array of separate subjects is not quite that large. Nevertheless, it is possible to identify users who are interested in chess for instance or the latest football scores.

### *Subject pathways*

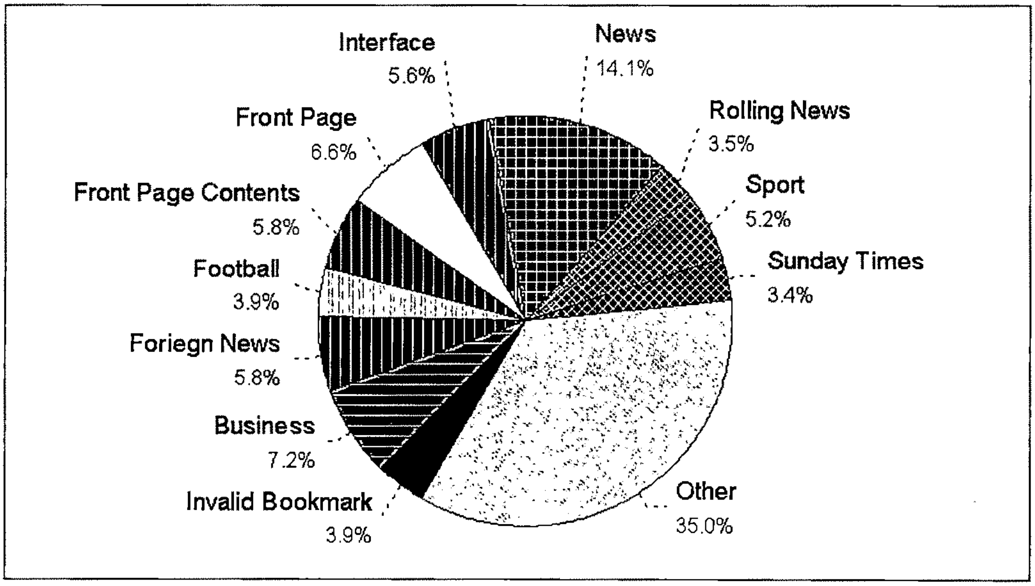
Users typically visit more than one page, so it is possible to determine what subject associations they make, which are the well trodden search pathways – and which are not. As important as knowing which pages/subjects are used is knowing which are not. Given the sheer size and complexity of most web sites, to have information that can show page obsolescence rates, enables pages to be pruned.

In fact it is possible, and highly illuminating, to classify users into types depending on entry point, navigation within and exit point. Interestingly, there is circumstantial evidence that users have different path behaviour between work and leisure browsing. The difficulty with a strictly behavioural approach is that each IP address reference may be used by a number of different users.

## Conclusion

There are two very important characteristics of the web that need to be borne in mind when evaluating web logs and comparing them with the logs of commercial online services and traditional library OPACs. Firstly, it is computers or computer networks that are the virtual users of the web. Log files of web servers record the action of these computers and computer networks – and not necessarily end-users. Secondly, the WWW system is essentially a system for the distribution of web pages.

*Figure 2. The Times web site: pages viewed*



Log files record the traffic in pages. The web was not constructed as an online searching service, although the presence of search engines give that impression and of course people use it as such. Therefore log files cannot be read in quite the same way as those of traditional library OPACs, for instance, and it can be particularly difficult to pin down use to an individual user.

This must be balanced by the fact that, quite unlike previous generations of computerised online systems (OPACs, CD-ROMs and online systems, for instance), the Internet is a genuinely popular service with a potentially unlimited reach. It is a draw for all types of people with, or without, an information need, many of whom will have no previous experience of searching computerised information systems. That is the attraction for information retrieval researchers, to attempt to discover what the general public, the country at large do – and do not do – when they go online and search and navigate themselves around similarly popular sources of information. The access logs and subscriber databases provide us with the detail in breathtaking volumes: but, not as they stand, all the answers. Even putting aside the problems of resolving use to a user, logs still do not tell us how appropriate the pages (information) were to the user or how satisfied they were with them. To obtain this data it is necessary to contact the user (via email, online questionnaire etc.) and ask them. Where sites have log-on facilities, and addresses are stored, this is a possibility, but could be regarded as intrusive or an offence under the Data Protection Act.

We all know that statistical data are far from problem free – and web transaction data are no exception. However, there is no doubt in our minds that with careful processing and interpretation of the data, and improved means of data capture, the quality of the data can be enhanced considerably.

## References

1. *ABC//Electronic. http://www.abc.org.uk*

2. Cohen, B. 'Below the surface of the web site figure.' *Electronic Mail & Guardian*, 18 November 1997. *http://mg.co.za/mg/news/97nov2/18nov-internet.html* [visited 22 January 1999]

3. Nicholas, D. 'An assessment of the online searching behaviour of practitioner end-users.' *Journal of Documentation* 52(3), September 1996, pp. 227-251.

4. Sandore, B. 'Applying the results of transaction log analysis.' *Library Hi Tech* 11(2), 1993, pp. 87-97.

5. Peters, T.A. 'The history and development of transaction log analysis.' *Library Hi Tech* 11(2), 1993, pp. 41-66.

6. Cutler, M. and Hall, D. 'Sizing 'em up.' *Internet World* 6(8), 1995, pp. 22-24.

7. Carter, D.S. Web server transaction logs, 1995. *http://www-personal.si.umich.edu/~superman/AP/Report.html#c5* [visited 18 January 1999]

8. Ottaviani, J.S. 'Archimedes: analysis of a HyperCard reference tool.' *College & Research Libraries* 56(2), 1995, pp. 171-182.

9. McGlamery, P. 'MAGIC transaction logs as measures of access, use and community.' *The Journal of Academic Librarianship*, November 1997, pp. 505-510.

10. Kurth, M. 'The limits and limitations of transaction log analysis.' *Library Hi Tech* 11(2), 1993, pp. 98-104.

11. IAB publishes guidelines for measurement of comparable online ad data. *http://www.iab.net/news/content/new%20/mmtf3pr.html* [visited 8 February 1999]

12. Keegan, V. 'What a web we weave.' *The Guardian G2*, 13 May 1997, pp. 2-3.

13. Wootton, C. 'Analyzing log files.' *Web Developer's Journal*, 8 November 1998. *http://www.webdevelopersjournal.com/articles/log_analysis.html* [visited 22 January 1999]

14. Fuller, R. and de Graffe, J. *Measuring user motivation from server log files. http://is.twi.tudelft.nl/~graaff/webuse96/* [visited 8 February 1999]

15. *How not to use HTTP log files. http://selsvr.stx.com/~eryq/using-http-logs* [visited 20 January 1999]

## Acknowledgements

The work on which this article was based was generously funded by News International.

# Information available on UK agricultural web sites

*Gillian M. Craig*
*Internet Studies Research Group, Department of Information Science,*
*City University, Northampton Square, London EC1 OHV*
*Email: craig@patrol.i-way.co.uk*

**Abstract** *In a survey of 132 UK agricultural web sites, more than 60% mounted directory-type information like that found in printed directories of information sources, but this was not always easily accessible. Information about people in organisations varied between the different categories of web site. Just under three-quarters of all sites mounted some substantive content but the amount of content on sites was small. Documents were available on slightly more than half the sites, but other types of content were available on fewer sites. Few sites explained how to locate information on the site or stated the intended audience. Less than one-third of sites had facilities to aid site navigation, but there were differences between the site categories for these features. Just over half the web sites displayed revision dates, but only about one-third had been updated in the last three months.*

## Introduction

As the quantity of information being made available via the Internet continues to escalate, increasing efforts are being made to select and classify 'quality' Internet resources so that they can be accessed more easily by users through subject-specific gateways[1]. In the UK the eLib Access to Network Resources Programme supports projects including OMNI (Organising Medical Networked Information), SOSIG (Social Science Information Gateway) and EEVL (Edinburgh Engineering Virtual Library), but there are many similar initiatives in other countries.

Agricultural Internet resources were amongst the first to be organised into listings and directories[2] and in the UK agriculture was included in the original BUBL Subject Tree Project in gopher in 1993[3]. At this stage the agricultural Internet resources available were primarily American. As early as 1995 a proposal for a Networked Agricultural Information Service in the UK was put forward by Stoker and Cooke[4] who highlighted recommendations made more than 15 years earlier in the report of the Agricultural Information Review Committee[5]. They suggested that a more integrated and coherent agricultural information system in the UK might be achieved through the establishment of a UK Internet site for agricultural information interests along the lines of the Bulletin Board for Libraries (BUBL) model. Unfortunately funding was not available for this project.

In October 1997 the state of development of networked agricultural information in the UK was unclear. There were a number of different directories and lists of links to UK agricultural web sites but these varied in their scope and coverage and none of the directories appeared to be comprehensive. The number of UK agricultural web sites listed in the BUBL LINK service, a multi-subject gateway, was still relatively small and, in the absence of any helpful published literature, it was not possible to establish accurately whether many quality UK agricultural resources were available on the web.

At the same time, countries such as the United States with its National Agricultural Library (NAL) and Agriculture Network Information Center (AgNIC)[6], Germany with its Agricultural Information Network, DAINet (Deutsches Agrarinformationsnetz)[7] and international organisations such as the Food and Agriculture Organisation of the United Nations (FAO) with its World Agricultural Information Centre (WAICENT)[8] were demonstrating a high level of commitment to the development of networked electronic agricultural information.

Part of the author's investigation carried out between November 1997 and September

1998 was to determine what UK web sites have to offer in terms of substantive quality agricultural information resources. Web sites accessible free of charge in all sectors of UK agriculture were included, but the focus was on research, education and extension in the core areas of crop and animal production.

UK agricultural web sites were identified using a variety of general and agricultural web indexes, directories and lists, and a total of 132 were selected for the investigation. These sites were placed in one of six categories: university sites (42), college sites (17), research sites (28), government and quasi-government sites (9), association and society sites (14) and commercial and miscellaneous sites (22). In the main survey in March 1998 each site was surveyed for 26 criteria designed to determine the different categories of information available on the web site, how easy it was to locate on the site and its currency. The second and third surveys in June and September 1998 were undertaken to identify changes in the quantity and categories of information available and its currency. In September 1998 a further 28 sites were also identified and surveyed.

## Accessibility of UK agricultural web sites

Although the total number of agricultural resources listed was still small, BUBL LINK pointed to 16 of the most significant UK web sites identified in the investigation and provided brief descriptions and metadata for them. It also provided links to some of the most useful agricultural web sites in other countries, for example, NAL's AgNIC in the United States, CAFFEINE: Canadian Agriculture Farm and Food Extension Information Network and Exchange, Wilfred Drew's classic 'Not Just Cows' guide, first produced in 1992, and the Agrigator Worldwide Agricultural Site Index maintained at the University of Florida. Two of the three UK agricultural site indexes used in the investigation, @grilink Online Rural Enterprise (Agrilink Directory) and AgScan 2000 were also included in BUBL LINK, but Jason Rankin's Agrilaunch was not.

In October 1998 on the new web site of the European Federation for Information Technology in Agriculture (EFITA) there was a Directory of agricultural sites in 20 European countries, but no entries for the UK were given. In the Regional Directories Section of NAL's AgNIC, AgScan 2000 was the only UK web site listed.

All but two of the UK agricultural web sites surveyed in March 1998 were still accessible in September 1998 and, with a few exceptions, information available on the UK agricultural web sites surveyed was free of charge. The Rural Business Network[9], linked from the ADAS web site, was available only to subscribers, as was Farming Online. There was also some information on the *Farmer's Weekly* Interactive site which was for subscribers to the Premium Service. The UPDATE database[10], accessible via the Welsh Institute of Rural Studies web site was also a subscription service, although a limited selection of the database was available for a free trial.

## Types of information available on UK agricultural web sites

### *(a) Directory-type information*

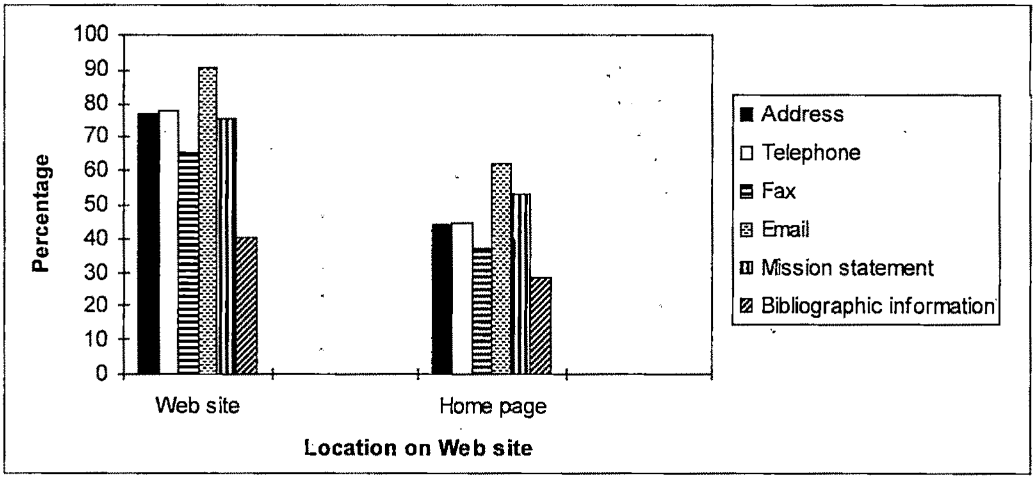
Directory-type information, the information about organisations usually found in traditional printed directories of information sources, was taken to include the address, telephone number, fax number and email address of the organisation, a mission statement and bibliographic information.

Although more than 60% of the surveyed web sites provided most of these items of directory-type information somewhere on the site, a smaller percentage mounted it on the home page or one mouse click from there (Figure 1). Over 70% of sites provided a mission statement, usually for the organisation rather than the web site, but only about 40% of sites provided any bibliographic information about the publications they produce. About 64% of association and society web sites and 57% of research sites announced publications on their web sites compared with 18% to 43% of other sites.

### *(b) Organisation information*

About one-third of the web sites surveyed gave clear information about the organisational infrastructure, naming and describing the activities of departments, sections and units within the main organisation. Many of the sites also

*Figure 1. Percentage of all agricultural Web sites mounting directory-type information, March 1998.*



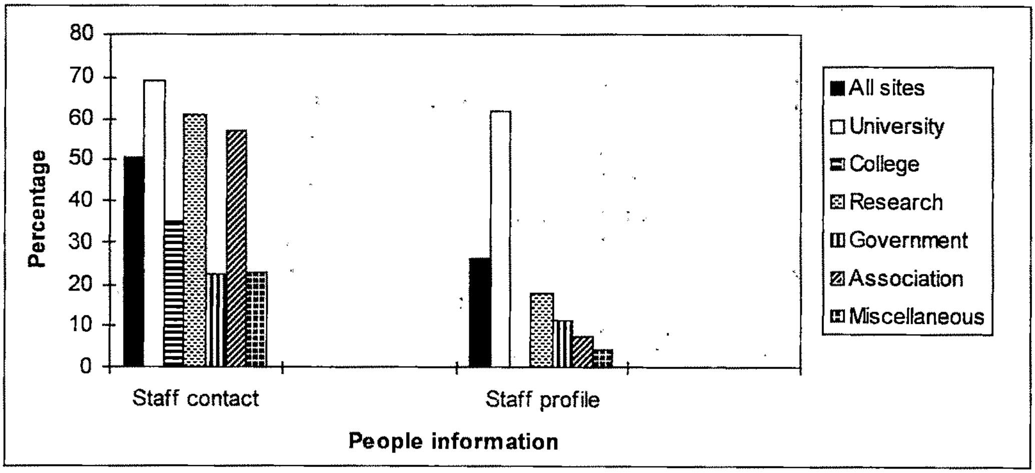
mounted a great deal of 'about' information of the sort usually found in annual reports, brochures and prospectuses.

### *(c) People information*

*Staff contact details*

Only half the web sites surveyed provided any type of contact information for some or all of the staff in the organisation, but the percentage of university, research and association web sites providing this information was notably higher than for the other categories of web site (Figure 2). Some sites mounted departmental lists of staff names with links to individual personal email forms. Other sites had search facilities for finding the contact details for staff in an organisation. This usually required knowledge of the name of the individual, so would be less suitable for the user wanting to locate an expert or subject specialist in a particular organisation without prior knowledge of their name. On other sites staff contact details were often provided in the form of internal telephone extensions.

*Figure 2. Percentage of different categories of UK agricultural Web sites mounting people information, March 1998.*



*Staff profiles*
More than 60% of university web sites mounted at least one detailed staff profile or personal web page compared with less than 20% of research web sites. There is considerable potential for mounting information about personal research activities and publications on personal web pages.

### *(d) Substantive online information resources*

In the surveys in March and September 1998 the agricultural web sites were monitored for items of substantive online information, often referred to as 'content'. These resources were placed in one of five categories: full texts of news items, press releases or newsletters; full texts of documents or fact sheets; data or statistical information; images or maps; and software.

In March 1998, 93 web sites (70.5%) had a minimum of one item of content compared with 89 web sites (69.5%) in September 1998. For each site category the number of sites with content was also very similar in March and September 1998. Some web sites mounted items of content in more than one of the five categories of specific information content. This put the overall total of sites mounting the different categories of content above 93 and 89 for the two survey dates respectively.

The most common two categories of substantive information were full texts of news and documents (Figure 3). However, in March 1998 these categories of information were available on only 37.1% and 46.2% of the web sites respectively. By September 1998 news and documents were available on slightly more web sites, 39.8% and 53.3% respectively. Data, images and software were only available on 5.3-12.9% of web sites in March, but their availability almost doubled over the six-month period.

Although around 70% of all the web sites surveyed mounted at least one item of substantive content, there remained more than one-quarter of sites surveyed that had nothing in the way of the content itemised earlier. When the different categories of substantive content were investigated separately in September 1998, only 52% of sites mounted at least one full text document, the most frequently occurring category of content. About 40% of sites mounted at least one full text news item, and only 23%, 15% and 13% of sites mounted any data, images or software. Although items of content were not counted individually for each site, it was clear that the number of items was small. It was possible, however, to single out some 38 sites with noteworthy items of content. A few of these sites were not as up-to-date in September 1998 as they had been at the start of the investigation.

Some patterns seemed to emerge between the different categories of web site mounting different types of resources. University web sites were the least unlikely to mount news and documents, while government web sites were most likely to exploit the web for the dissemination of their news and press releases.

### *(e) External hypertext links*

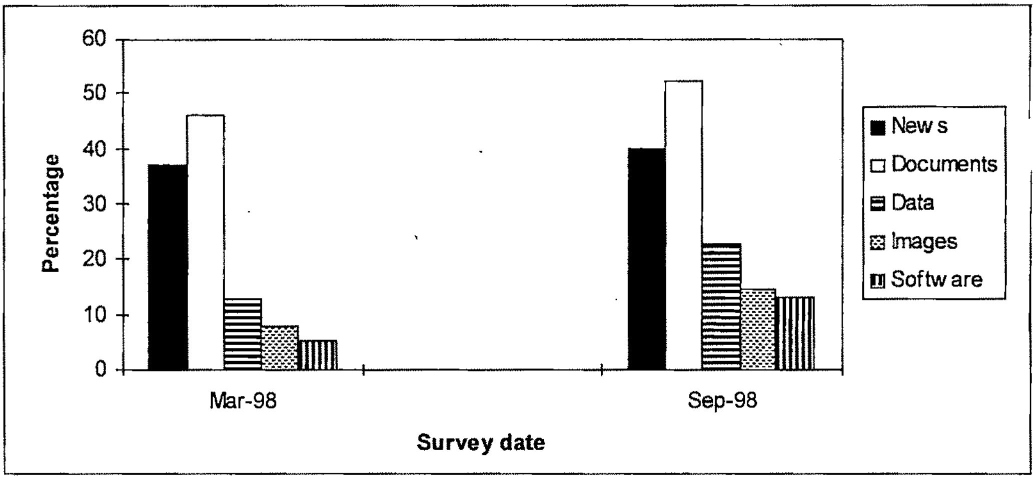
In March 1998, just over 60% of all web sites had external links to related web sites (Figure 4). University and government web sites were the least likely to have external links and the miscellaneous web sites had the highest percentage of sites with links. This category included several new sites developed as Internet guides and consisting primarily of site links. These included Ukexnet Horticultural Index, @grilink, Agnet, Agrilaunch and Pigs UK. In this category the Galaxy Precision Ag services site had good precision farming and remote sensing links, dotFARMING had lots of links arranged in nine sections and including livestock, cropping, machinery and agribusiness, and also an 'all about farming links' section.

The Wye College Department of Agricultural Economics and Management had hyperlinks for Central and Eastern European agricultural information and the Royal Agricultural College, Sparsholt College, Merristwood College and Greenmount College all had large numbers of links, arranged in subject sections.

### *(f) Events*

In March 1998 just under 40% of all sites provided information on events of some sort, and association web sites were the most likely sites

*Figure 3. Percentage of all UK agricultural web sites mounting at least one item of different categories of online content, March and September 1998.*



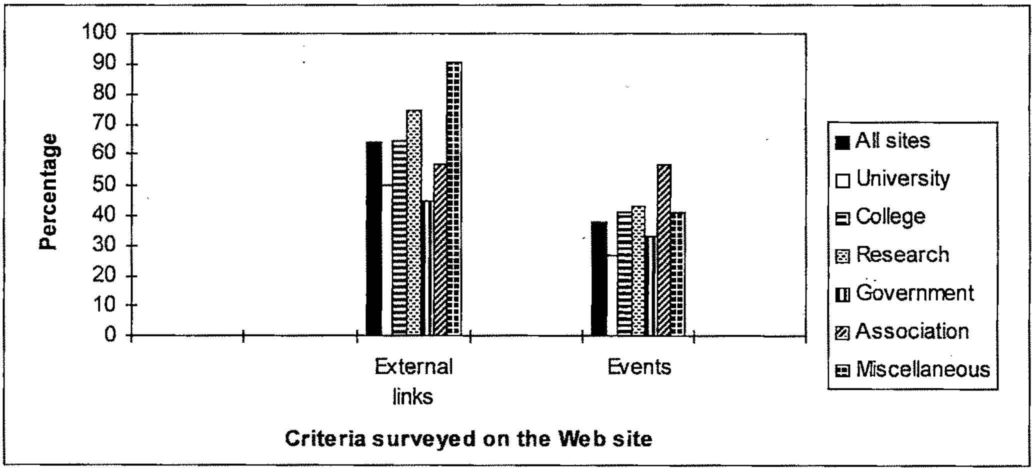
to mount information of this kind. Just under 60% of association web sites mounted information about events compared with just over 25% of university web sites and around 40% of the other categories of web site (Figure 4).

In November 1998 the Royal Agricultural Society was due to launch its new 90 page Royal Show web site, a development which was seen 'as an important step towards keeping both visitors and exhibitors informed'[11].

## Locating information on UK agricultural web sites

Surprisingly very few of the web sites surveyed provided a textual description of the site, its contents, intended audience and how best to locate information on the site. A good example of this feature, however, was the 'About' information mounted on the Plant Pathology Internet Guide Book site, mirrored on the web site of the British Plant Pathology Society. Sev-

*Figure 4. Percentage of different categories of agricultural web sites with external links and information about events, March 1998.*



eral web pages of text covered the mission of the Guide Book, how to use it, its features, its author and 'your feedback'.

Of the UK sites sampled, the most useful guidance on using the site was found on the welcome page of the Greenmount College site. This site also mounted weekly lists of resources added to the site, an A-Z index of contents and an online Internet Training Manual. Other UK sites mounting useful features to help site navigation included the MAFF site, the Southampton School of Biological Sciences site, the Cyanamid agriCentre site of Cyanamid Agriculture UK, Bunn's Fertilisers Online and Galaxy Precision Ag Services.

The number of links from the home page was very variable ranging from one click to enter the site to as many as 60 on Jason Rankin's Dairy Page. However, the majority of sites (65.9%) had between 7 and 19 links from the home page, 9.8% of sites had 1-6 links and 24.2% had 20-60 links.

### *(a) Search facility*

In March 1998 just over 25% of all sites had some sort of search facility. Nearer 40% of research and government web sites had a search facility compared with 20-25% for the other categories of site (Figure 5).

By September 1998 over 40% of research and government web sites displayed a search facility on the home page and there was a marked increase in the percentage of miscellaneous web sites, also to 40%. University, college and association web sites were least likely to have a search facility.

### *(b) What's new*

Although almost 20% of all sites had a 'What's new' facility in March 1998, a much higher percentage of government and miscellaneous sites (33-36%) and a much lower percentage of other sites (about 6-17%) had such a facility (Figure 6). For college sites there was little change over the six month period to September 1998, for research there were slightly more sites and for universities there were fewer sites with a'What's new'section. Government sites were unchanged at a higher level while the percentage of association and miscellaneous web sites with 'What's new' increased notably.

### *(c) Help*

In March 1998 the percentage of sites with 'Help' was just under 8% for all sites and was low for all categories of site except government sites where the percentage was slightly higher at just over 20% (Figure 7). This changed a little by September in that the percentage of government sites with 'Help' increased appreciably to over 30%. The percentage of miscellaneous sites with 'Help' also increased but from just under 10% to 15%

*Figure 5. Percentage of different categories of UK agricultural web sites with a search facility, March and September 1998.*

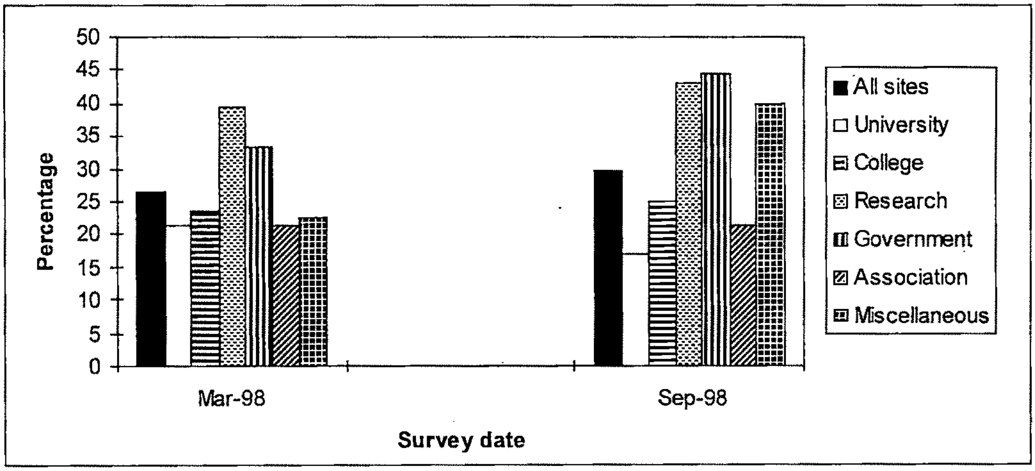
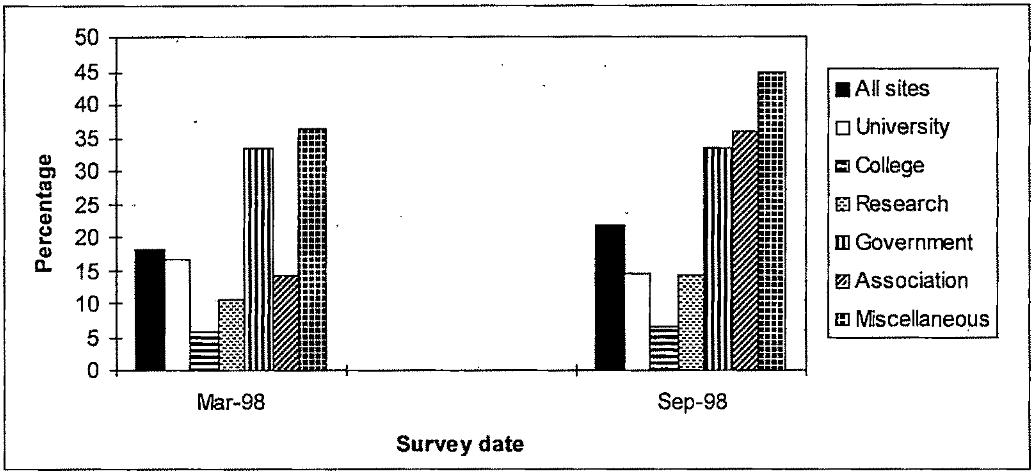


*Figure 6. Percentage of different categories of UK agricultural web sites with a 'What's new' facility.*

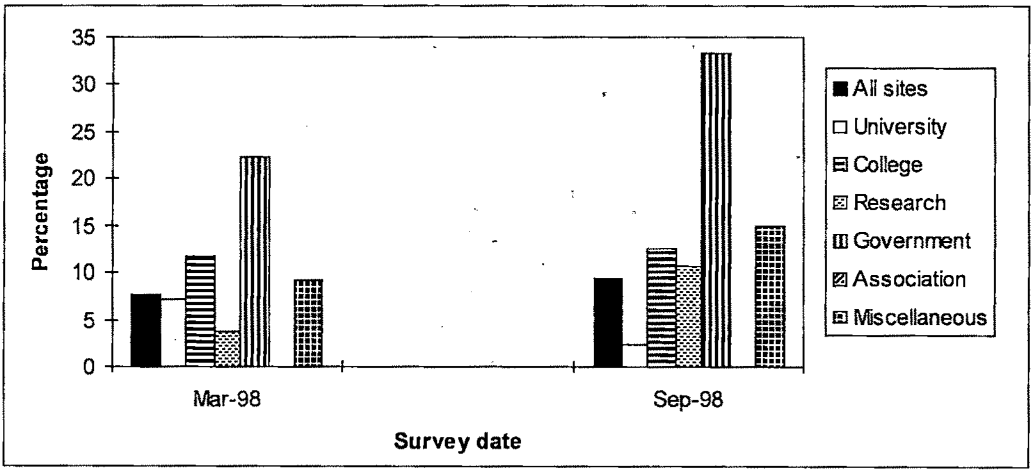


and the percentage of research sites increased from just under 4% to 10%. There were no 'Help' facilities on association web sites.

**Currency of UK agricultural web sites**

In March 1998, 61 sites (46.2% of those surveyed) displayed a revision date and all but three of these sites gave this information on the home page (Figure 8). However, it was not always clear whether this date of revision referred only to the home page or to the whole site, and whether revision dates were always changed rigorously when new material was added to the site. Some sites had revision dates for individual pages of information on the site. Of the 61 sites which were displaying revision dates in the March 1998 survey, 13 sites had been updated in February or March 1998, 14 in January 1998, 21 in the months between July and December 1997 inclusive, six between January and June 1997 inclusive and five in 1996. Almost 60% of university sites displayed revision dates compared with 35-45% for other categories of site.

*Figure 7. Percentage of different categories of UK agricultural web sites with a 'Help' facility, March and September 1998.*



When the 132 web sites were surveyed for the second time in June 1998, 33 sites (25%) displayed revision dates (either on the home page or on specific new items of information) which indicated that they had been updated within the last three months and 13 sites had been updated that week (7 to 13 June 1998). For 25 sites (18.9%) the revision dates displayed during the March survey were unchanged and many of these appeared not to have been updated for over six months.

By September 1998 more sites displayed revision dates. The greatest increases were in the percentages of miscellaneous sites followed by government and research sites. Of the 69 sites (53.9%) with revision dates, 45 sites (35.2% of the total surveyed) had been updated within the last three months, nine (7.03%) three to six months ago and 15 (11.7%) more than six months ago. Nobody would dream of issuing a print publication without a date but it appears still acceptable to provide web sites without any information by which to judge the currency of their contents.

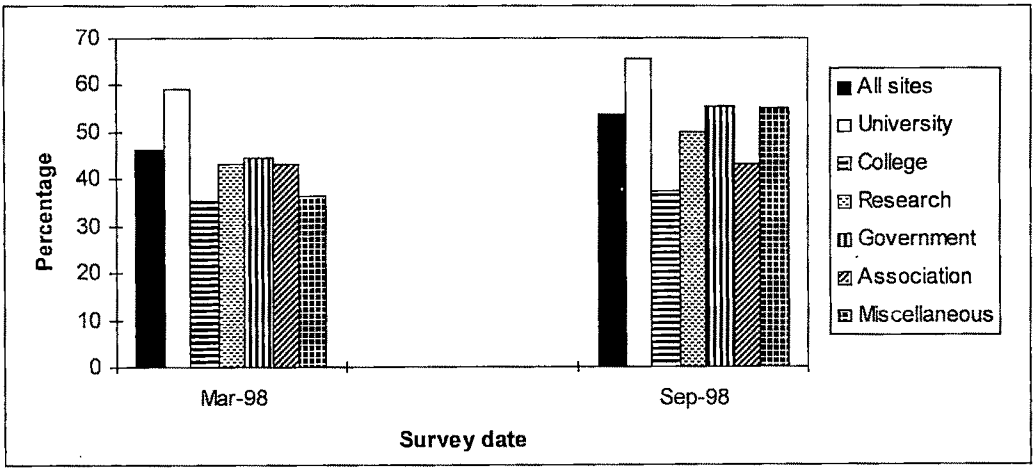
## Information from printed directories of information sources

A comparison was made between the directory-type information on the web sites surveyed and that contained in two printed directories, the *Aslib Directory of Information Sources in the United Kingdom*[12] and the *Directory of Rural Organisations*[13]. Both printed directories had many more entries for UK agricultural organisations than the number of web sites identified and surveyed in March 1998. However, only 56 of the entries in the *Aslib Directory* and 75 of the entries in the *Directory of Rural Organisations* had a corresponding web site in the March 1998 survey. The 9th Edition of the *Aslib Directory* was the first to include Internet resources and 24 of the 56 *Aslib* entries which matched web sites surveyed included an Internet address. This compared with only seven of the entries in the *Directory of Rural Organisations.*

Information made available from the 10th Edition of the *Aslib Directory* prior to its publication (Reynard, personal communication) indicated that there were about 13 more directory entries matching the web sites surveyed in March 1998, but less than half of the directory entries with Internet addresses gave details of the type of information on the site and even fewer gave pointers to substantive information content.

Although the web sites of agricultural organisations were potentially able to provide a much greater quantity and range of information than that contained in printed directories, some sites did not provide all the items of directory-type information contained in the printed directories, and others did not present the information in a clear and structured man-

*Figure 8. Percentage of different categories of UK agricultural web sites with a revision date, March and September 1998.*



ner as in the printed directories. However, in the absence of a gateway for UK agriculture Internet sites, a printed directory with Internet addresses for all entries, where available, could be one useful starting point for locating the Internet sites of UK agricultural organisations.

## Discussion and conclusions

Now that a significant number of UK agricultural organisations have a presence on the World Wide Web there are some important questions to be asked. How do you find these sites quickly and easily? What useful information might be available on them? Who is it aimed at? Is it easy to locate once you have located the site? Is it up-to-date? This investigation starts to answer these questions, but they should also be addressed by all those in the UK who are involved with the dissemination of agricultural information and with the development and maintenance of UK agricultural web sites.

Anyone who uses the WWW frequently soon assembles a personal list of 'bookmarks' or 'favourites' which can be used to speed up access to useful web sites. Internet subject directories and subject gateways to quality Internet resources also assist the user to access potentially useful information more quickly than with search engines.

In the absence of a major UK gateway for quality agriculture Internet resources, directories such as @grilink Online Rural Enterprise, the AgScan 2000 Search Index for UK Agriculture and Agrilaunch are useful starting points for UK resources, and BUBL LINK also points to some of the most useful UK agricultural web sites.

However, there is also a place for a comprehensive and up-to-date list of UK agricultural web sites arranged by organisation. Two such lists, used at the start of the present investigation, have not been developed and kept up to date, but in the work outlined in this paper a new list of 160 UK agricultural web sites and their URLs, arranged by category, has been created. This is an additional resource to be made available.

There appeared to be an enormous variation both in the amount and range of agricultural information mounted on UK web sites and the ease or otherwise with which significant items of information could be found on the sites. There were some notable web sites, such as Greenmount College, RIO - Roslin Institute Online, the Ministry of Agriculture, Fisheries and Food, and CLUES (Centre for Computer-based Learning in Land Use and Environmental Sciences), which demonstrated commitment and enthusiasm for disseminating agricultural information via the web. For other web sites it appeared that, in a rush to establish an Internet presence, organisations mounted web pages without clearly defining their intended audience and then failed to develop and improve their sites.

The total amount of substantive online information available on UK agricultural web sites was very small, especially when compared with that in the United States where networked agricultural information is especially well developed. It has been a long-standing mission of the USDA to encourage agricultural research and to disseminate information. Accordingly, over the years significant financial resources have been allocated by the United States Government for agricultural information R&D. The NAL has also been a leading light in the transition to electronic information and, most notably, has developed the AgNIC web site[14] from which a whole range of value-added services, including a number of directories for agriculture related information systems, databases and datasets, experts in agriculture, agriculture resources available over the Internet, an agricultural conferences, meeting and seminars calendar, an online reference service, educational services and liaison activities are available.

In the UK, agricultural research, education and extension are not all under a national department of agriculture umbrella and there is no tree-like organisational infrastructure and national agricultural library as in the United States. The agricultural situation in the UK has changed enormously in the late 1980s and 1990s with major reductions in state funding for specific areas of agricultural research and extension and changes in the way further and higher education are funded. These changes, have been far-reaching and are ongoing[3,15,16,17].

In the agriculture field where different levels of information user from scientific researchers, through educators and extension specialists to farmers have been recognised for many years[5,18,19,20] it is particularly important to mount Internet resources which are designed for specific users and to indicate this on the site. Those involved with agricultural information dissemination at whatever level must also remember that British farmers are the ultimate end-users of information emanating from research and development. The process of communicating and reformulating relevant scientific information so that it is helpful to the farmer and can be applied where appropriate to farming businesses continues to be a thought-provoking subject[20,21].

Chamberlain[15] has stressed the role played by the British agricultural press in communicating technical material to farmers and it is significant that web sites such as *Farmer's Weekly* Interactive (Fwi) and dotFARMING, the web site of the Miller Freeman magazines, were among the most notable and valuable found in the survey. Chamberlain also stressed the uptake by farmers of the subscription services like Farming Online and the Rural Business Network.

The quantity of networked agricultural information available worldwide will undoubtedly increase in the future and as other countries are seen to be making efforts to improve their national information systems and services it seems clear that the UK as a country should be doing the same. However, to date, developments in the UK seem to have been confined to a relatively small number of individual agricultural organisations and enterprises.

Four years ago Craig[3] made a series of recommendations following a review of agricultural and land-use information in the UK. They included keeping abreast of agricultural information work at a European and International level, funding for specific information projects in agricultural subject areas, better 'sign posting' of information resources in traditional and electronic formats and a national agricultural information conference to provide a new focal point for UK agricultural interests.

Very little has happened until recently but October 1998 marked the Inaugural Conference of BAITA (the British Association of IT in Agriculture). BAITA is dedicated to the development and advancement of the use of IT in the agricultural industry and it brings together people from all parts of the industry: farmers, growers, consultants, educationalists, sales, IT specialists and research workers. BAITA is affiliated to the European Federation for IT in Agriculture (EFITA) and this Federation had its first conference in Copenhagen in June 1997.

Hopefully, with its web site, newsletter and conference, BAITA will attract members of the library and information science profession and help to improve the communication of agricultural information and agricultural information issues in the UK.

## References

1. Bond, D. and MacLeod, R. *Pinakes: a subject launch pad.* (Edinburgh: Heriot-Watt University, 1998.) *http://www.hw.ac.uk/lib/WWW/irn/pinakes/pinakes.html*
2. Besemer, H. and Veerman, I. 'Agricultural information on the Internet: what is out there and how to find it.' *IAALD Quarterly Bulletin* 40 (2/3), 1995, pp. 61-67.
3. Craig, G.M. *Agricultural and rural land-use information in the UK.* British Library R&D Report 6189. (London: British Library Research and Development Department, 1995.)
4. Stoker, D. and Cooke, A. 'Using the Internet to co-ordinate the provision of agricultural information in the UK.' *IAALD Quarterly Bulletin* 41(2), 1996, pp. 196-200.
5. Craig, G.M. *Information systems in UK agriculture. Final Report of the Agriculture Information Review Committee.* British Library R&D Report 5469. (London: British Library Research and Development Department, 1979.)
6. Thompson, R.E. 'Agricultural Network Information Centre (AgNIC).' *Agricultural Libraries and Information Notes* 22 (1-3), 1996, pp. 6-10.
7. Freidrich, H. and Pohlman, J.M. 'The German Agricultural Information Network Deutsches Agrarinformationsnetz (DAINet) – goals, state of the art and future perspectives.' *IAALD Quarterly Bulletin* 42 (3/4), 1997, p. 146.
8. Judy, J.R., Mangstl, A. and Le Hunte Ward, F. 'WAICENT, World Agricultural Information Cen-

tre: FAO's information gateway.' *IAALD Quarterly Bulletin* 42 (3/4), 1997, pp. 136-145.

9. Offer, A. *The rural business network in the UK.* (Copenhagen, Denmark: The Royal Veterinary College and Agricultural University, 1998.) *http://www.efita.org/ conference97/paper.asp*

10. Evans, A. UPDATE - the farming and countryside index. Unpublished paper submitted to the Joint GBDL, IAALD, CTA, NITRANET Conference, 2-5 June 1998. (Freising, Germany: GBDL, 1998.)

11. Royal Agricultural Society of England. 'Royal Show joins the web.' *News: the Newsletter for Members of the Royal Agricultural Society of England* 8, 1998.

12. Reynard, K.W. and Reynard, J.M.E. (eds.). *Aslib Directory of Information Sources in the United Kingdom.* (9th Edition) (London: Aslib, 1996.)

13. Powell, E. and Taylor, D. *Directory of Rural Organisations.* (Ipswich: Farming Press, 1997.)

14. Thompson, R.E. *Managing electronic information in agriculture. Toward an agriculture network information center. A model for distributed access. Summary report of AgNIC Planning Workshop December 4, 5 & 6, 1994 Washington, D.C.* (Beltsville, USA: National Agricultural Library, 1995.) *http://www.agnic.org/agnic/ agnic.html*

15. Chamberlain, D. 'Dissemination of information in agriculture.' *Journal of the Royal Agricultural Society of England* 158, 1997, pp. 8-13.

16. Norman, L. 'Whither agricultural education?' *Journal of the Royal Agricultural Society of England* 159, 1998, pp. 158-166.

17. Jones, E.W. 'Contemporary issues shaping the future of English higher education in agriculture.' *Journal of the Royal Agricultural Society of England* 159, 1998, pp. 167-183.

18. Maltha, D.J. 'Information needs in agriculture.' *IAALD Quarterly Bulletin* 10(2), 1972, pp.170-177.

19. Brennan, P.W. Information flow in American agricultural literature. *IAALD Quarterly Bulletin* 20(2), 1995, pp. 86-93.

20. Gasson, R. and Hill, B. 'Technology interaction in agriculture and horticulture.' *Journal of the Agricultural Society of England* 159, 1998, pp. 148-157.

21. Bunyan, P., Bailey, A. and Buckenhan, S. 'Technology transfer in the food and agriculture industries.' *Journal of the Royal Agricultural Society of England* 156, 1995, pp. 62-68.

## UK web sites mentioned in the text

- ADAS *http://www.adas.co.uk*
- @grilink Online Rural Enterprise *http://www.agrilink.co.uk/*
- Agnet (includes AgScan 2000 Search Index and Busitool) *http://www.agnet.co.uk/agnet/agnet. htm*
- Agrilaunch *http://www.agrilaunch.com/*
- British Association for IT in Agriculture (BAITA) *http://www.baita.co.uk/*
- British Society for Plant Pathology *http://www.bspp.org.uk/*
- BUBL LINK *http://bubl.ac.uk/link/*
- Bunn's Fertilisers Online *http://www.jhbunn.co.uk/*
- CLUES Centre for Computer-based Learning in Land Use and Environmental Sciences *http://www.clues.abdn.ac.uk:8080/*
- Cyanamid agriCentre *http://www.agricentre.co.uk/*
- DotFARMING – Your gateway to agriculture on the Internet *http://www.dotfarming.com/*
- Edinburgh Engineering Virtual Library (EEVL) *http://eev.ac.uk/*
- Farmer's Weekly Interactive *http://www.fwi.co.uk/*
- Farming Online *http://www.farmline.com/*
- Galaxy Precision Ag Services *http://www.jhbunn.co.uk/galaxy/*
- Greenmount College of Agriculture and Horticulture *http://www.greenmount.ac.uk/*
- Jason Rankin's Dairy Page *http://www.loughries. demon.co.uk/index.htm*
- Merristwood College *http://www.merristwood.ac.uk/merristwood/*
- Ministry of Agriculture Fisheries and Food, UK *http://www.maff.gov.uk/*
- Organising Medical Networked Information (OMNI) *http://omni.ac.uk*
- Pigs UK *http://www.pigsuk.com*
- Roslin Institute *http://www.ri.bbsrc.ac.uk/*
- Royal Agricultural College *http://www.royagcollege.ac.uk/*

- Royal Show *http://www.royalshow.org.uk/*
- Social Science Information Gateway (SOSIG) *http://sosig.ac.uk/*
- Southampton University School of Biological Sciences *http://www.sobs.soton.ac.uk/sobs/*
- Sparsholt College *http://www.sparsholt.ac.uk/*
- Ukexnet Horticultural Index *http://www.ukexnet.co.uk/hort/*
- Welsh Institute of Rural Studies *http://www.aber.ac.uk/~wirwww/*
- Wye College Department of Agricultural Economics and Business Management *http://www.wye ac.uk/Agecon/index.html*
- Agriculture Network Information Center (AgNIC) *http://www.agnic.org/*

**Other web sites**

- Agrigator Worldwide Agricultural Site Index *http://gnv.ifas.ufl.edu/www/agator/htm/ag.htm*
- CAFFEINE: Canadian Agriculture Farm and Food Extension Information Network and Exchange *http://www.caffeine.ca/index.cfm*
- DAINet *http://www.dainet.de*
- European Federation for Information Technology in Agriculture (EFITA) *http://www.efita.org*
- Not Just Cows *http://www.snymor.edu/~drewwe/njc*
- Plant Pathology Internet Guide Book *http://www.ifgb.uni-hannover.de/extern/ppigb/ppigb.htm*
- World Agricultural Information Centre (WAICENT) *http://www.fao.org/WAICENT*

# On universality in productive and information technology

*Julian Warner*
*School of Management, The Queen's University of Belfast, Belfast BT7 1NN*
*Email: j.warner@qub.ac.uk*

**Abstract** *Information science has differentiated information technology from productive technology, but the common concept of technology remains largely unexplored. A view of technology as a human construction, applied to productive technology, has begun to be developed to comprehend information technology. Information technology is regarded as a form of knowledge concerned with the transformation of signals from one form or medium into another. Analogous, although not identical, concepts of universality can be distinguished for both productive technology and for information technology. The steam engine has been regarded as a universal source of motive power and of the computer as a universal information machine. Universality helps account for the wide adoption of the steam engine and the computer. For both forms of technology, theoretical considerations related to universality and working constructions embodying universality have, in contrasting ways, been partly separate developments. Further possibilities offered by a view of information technologies as human constructions are indicated.*

## Introduction

Information science has tended to exclude technologies concerned with the human transformation of the material environment (productive technologies) from its scope and has preferred to focus on information technologies and their applications. Where explicit attention has been given to productive technologies, for instance in discussions of computability and of the computer as a machine[1], the primary, if not exclusive, concern has been to differentiate information technology from productive technology. Information technology is then regarded as concerned with the transformation of information inputs into information outputs rather than, for instance, the production of power from potential energy. More commonly, a disjunction between information and productive technology is implicit, for instance in the antithetical distinction of information from industrial societies[2]. The differentiation of information from productive technology is, then, clearly established but the common concept of technology remains unexplored.

The lack of examination of technology has limited the theoretical sophistication brought to understanding information technology by information science and there have been other deficiencies. Historical perspectives have been restricted. Elements of technological determinism, also present in lay thinking, have been identified and criticised[3]. Yet strong traces still remain, most subtly and pervasively in the limited recognition of information technology as a human construction. Developments in information technology can to be treated as if they were autonomous developments, particularly by the language used, for instance, by references to the evolution of technology, although this implied position need not be fully articulated[4,5]. The possibility of regarding technologies as human constructions, invented and adopted under specific historical conditions, is seldom developed. The technological determinism is also shared with some discussions of productive technology, where it has also been criticised[6,7].

Some recent developments have run counter to this lack of theoretical reflexivity. A discipline of the history of information, which would subsume the history of information services and of libraries, has been proposed[8]. The developing field of social informatics includes the social shaping of technology as one of its foci[9]. A view of technology as a human construction, derived from Marx (with some curious connections to Babbage), has begun to be adapted to comprehend information technology[10,11]. Marx's analysis of productive technology has been regarded as sophisticated, little used, but yielding high marginal returns when applied[12,13,14] and its application to information technology could be comparably fertile.

From the basis offered by Marx, a modern understanding of productive technology as a special form of knowledge dealing with the transformation of the material environment into useful goods and services has been developed, particularly by Rosenberg[15,16,17]. Some further critical distinctions are made: invention is distinguished from innovation, or the adoption of the invented product, and diffusion mechanisms and path-dependency, the strong influence of early events on subsequent developments, are recognised. On this model, information technology is regarded as a form of knowledge concerned with the transformation of signals from one form or medium into another, with the other crucial distinctions retained[10]. The distinction of innovation from invention can be particularly valuable and has been used in supporting indications that the late nineteenth century United States was a critical period for innovation in technologies since comprehended by the generic term, information technology, such as the telegraph and telephone[10,11]. The concern in this paper is not with the full development of this understanding of information technology but with indicating a specific analogy between productive and information technology, in the concept of universality.

## Concept of universality

A strong and revealing analogy, not a strict identity, will be indicated with regard to the concept of universality for productive and information technology. For both forms of technology, practical understanding and theoretical articulation can develop separately and working technologies produced on the basis of practical understanding can precede their full theoretical description.

### *Productive technology*

For productive technology, the steam engine constituted a universal source of motive power which could be substituted for human physical and animal labour and for natural sources of power adapted to human use, such as rivers.

> 'The greatness of Watt's genius showed itself in the specification of the patent that he took out in April 1784. In that specification his steam-engine is described, not as an invention for a specific purpose, but as an agent universally applicable in industry. Many of the applications he points out in it, for instance the steam-hammer, were not introduced until half a century later. Even so he doubted the applicability of steam to navigation.'[18]

The contrast made between the invention and its subsequent applications can be regarded as an exemplification of the difficulty of predicting the purposes to which inventions will be adapted. Although, in this passage, attention is given to an individual's contribution, Marx saw technology as a social process. At its simplest, the history of inventions is not the history of inventors[19].

In *Capital*, the adoption of the steam engine to replace human power is regarded as favouring large-scale industry over domestic or small-scale manufacture, by reducing the comparative costs of production[20]. It also reduces the dependence of industry upon particular locations, by substituting for power derived from rivers and, later as a component of transport technologies, enabling the greater movement on mineral resources[21]. From a perspective informed by the presence of the computer as a universal information machine, the universality of the steam engine as a source of motive power can be seen to account for its wide adoption.

The practical understanding embodied in working technologies, such as the steam engine, is regarded as a significant influence for the development of its theoretical articulation:

> 'Its [large-scale industry's] principle, which is to view each process of production in and for itself, and to resolve it into its constituent elements without looking first at the ability of the human hand to perform the new processes,

brought into existence the whole of the modern science of technology.'[22]

The development of practical understanding before theoretical articulation is reminiscent of the account of the origins of geometry in the need to recover field boundaries obliterated by the annual inundation of the Nile[23,24]. More familiarly, the method of thought described is similar to Cartesian methods of analysis and to Russell's logical atomism. The science of technology then

> 'discovered the few grand fundamental forms of motion which, despite all the diversity of the instruments used, apply necessarily to every productive action of the human body, just as the science of mechanics is not misled by the immense complication of modern machinery into viewing this as anything other than the constant re-appearance of the same mechanical processes.'[25]

Here there is a strong analogy with the isolation of the fundamental operations of mathematics, as the writing, erasure and substitution of symbols, and their, concurrent and subsequent, mechanisation.

In summary, then, universality can be distinguished in productive technologies and is exemplified by the steam engine. Its universality as a source of motive power helps explain its wide adoption. For productive technology, practical understanding precedes and stimulates scientific description and analysis. There are also emerging analogies with information technology, particularly in the method of analysis, the reduction of apparent diversity to a few fundamental operations, and in the concept of universality itself.

### *Information technology*

For information technology, theoretical considerations related to universality and practical constructions embodying universality have also been separate developments, to some extent. In contrast to productive technology, theory has partly preceded practice, particularly with regard to the computer as a universal information machine and here the effects of the late nineteenth century convergence of science and technology can be traced.

At a theoretical level, information theory as represented by *The mathemaical theory of communication* was concerned with the economic and accurate transmission of symbols at a theoretical level 'so general that one does not need to say what kinds of symbols are being considered'[26]. The reduction of operations in mathematics to the erasure, writing and substitution of symbols according to given rules occurred in the 1920s[27] and preceded the formulation of a universal information machine by Turing in 1936. Symbolic logic and automata theory are still not fully integrated with each other[24].

For practical constructions, codes for the transmission of messages preceded information theory and were discovered to be highly efficient when assessed against the theoretical possibilities for economic coding. Binary coding itself has a long history and could also be regarded from a semiotic perspective as exemplifying the minimal difference needed for signification[24,28]. The computer as a universal information machine is usually regarded as a separate development from automata theory[24,29], although other studies have qualified this and suspended judgement in advance of a careful study of the relevant intellectual history[30].[i]

The strict concept of universality applied to the computer is better known than its less highly formalised analogues in productive technology or information theory and codlings, although still not necessarily widely diffused or well understood. Even those highly familiar with the theoretical concept did not necessarily realise its full range of theoretical applications:

> 'If it should turn out that the basic logics of a machine designed for the numerical solution of differential equations coincide with the logics of a machine intended to make bills for a department store, I would regard this as the most amazing coincidence that I have ever encountered.'[31]

From a historical perspective informed by the multiplicity of computer applications, the universality of the computer can be seen to explain its diffusion, with the intellectual labour of programming displacing the intellectual and physical labour involved in the construction of special purpose information machines.

## Conclusion

A view of both productive and information technologies as human constructions has helped reveal one specific parallel between them. Analogous, although not identical, senses of universality were indicated. Universality of motive power and of signal transformation provided explanations for the wide diffusion of the steam engine and of the computer.

Further possibilities are offered by the understanding of information technology indicated. For instance, in contrast to productive technology, information technologies are not necessarily worn out by use (consider the erosion of the spindle in spinning)[32]. The rate of innovation in computer and other electronic products might then be explained by the need to stimulate purchases before replacements are required. In common with productive technology, the knowledge and labour embodied in the machine may reduce the human labour required for current processes: 'what else does growing productive power of labour mean than that less immediate labour is required to create a greater product ... ?'[33]. For example, a printed text can be produced with less labour at the point of production than that required for setting from moveable type. The ability of the computer to be programmed to undertake the primitive operations of writing, erasure and substitution of symbols is congruent with the reduction of employment in routine information handling occupations[34] .

A further possibility would be the full development of the approach implied by Walter Benjamin's remark that, 'within the phenomenon [of the possibilities of reproduction of art objects] which we are here examining from the perspective of world history, print is merely a special, though particularly important, case'[35]. For instance, to give full primacy to the economic and cultural over the technological, the adoption of printing and the wider dissemination of texts can itself be seen as a product of the stress on personal conscience in Protestantism. The influence of the computer as a universal machine, capable of combining functions previously distributed across different technologies, may become one factor in encouraging the recognition of disparate contemporary and historical forms as variants of information technology, as both universality and information technology begin to be better understood.

## References

1. Minsky, M. *Computation: finite and infinite machines.* (Englewood Cliffs, NJ: Prentice-Hall, 1967, pp.1-6.)

2. Bell, D. 'The social framework of the information society.' In: Forester, T. (ed.) *The microelectronics revolution: the complete guide to the new technology and its impact on society.* (Oxford: Basil Blackwell, 1980.)

3. Mackenzie, D. *Knowing machines: essays on technical change.* (Cambridge, MA: MIT Press, 1996, p.5.)

4. Kahin, B. and Nesson, C. (eds.) *Borders in cyberspace: information policy and the global information infrastructure.* (Cambridge, MA: MIT Press,1997.)

5. Warner, J. Review of Kahin, B. and Nesson, C. (eds.) 'Borders in cyberspace: information policy and the global information infrastructure.' *Journal of the American Society for Information Science* 49 (4), pp.387-388.

6. Mackenzie, D. *Knowing machines: essays on technical change.* (Cambridge, MA: MIT Press, 1996, pp.24-25.)

7. Rosenberg, N. *Inside the black box: technology and economics* (Cambridge: Cambridge University Press, 1982, p.35.)

8. Black, A. 'Information and modernity: the history of information and the eclipse of library history.' *Library History* 14(1), 1998, pp.441-45.

9. Kling, R., Rosenbaum, H. and Hert C. 'Social informatics in information science: an introduction.' *Journal of the American Society for Information Science* 49(12), 1998, pp.1047-1052.

10. Warner, J. 'An information view of history.' *Journal of the American Society for Information Science* (50), 1999 (in press).

11. Warner, J. 'Information society or cash nexus? A study of the United States as a copyright haven.' *Journal of the American Society for Information Science* (50), 1999 (in press).

12. Mackenzie, D. *Knowing machines: essays on technical change.* (Cambridge, MA: MIT Press, 1996, p.19.)

13. Rosenberg, N. *Inside the black box: technology and economics.* (Cambridge: Cambridge University Press, 1982, p.viii.)

14. Rosenberg, N. *Exploring the black box: technology, economics and history.* (Cambridge: Cambridge University Press, 1994, p14.)

15. Rosenberg, N. *Inside the black box: technology and economics.* (Cambridge: Cambridge University Press, 1982.)

16. Rosenberg, N. *Exploring the black box: technology, economics and history.* (Cambridge: Cambridge University Press, 1994.)

17. Rosenberg, N. *Perspectives on technology.* (Cambridge: Cambridge University Press, 1976.)

18. Marx, K. *Capital: a critique of political economy.* Volume One. Introduced by Mandel, E. Translated by Fowkes, B. (Harmondsworth: Penguin Books in association with *New Left Review*, 1976, p.499.)

19. Rosenberg, N. *Inside the black box: technology and economics.* (Cambridge: Cambridge University Press, 1982, p.48.)

20. Marx, K. *Capital: a critique of political economy.* Volume One. Introduced by Mandel, E. Translated by Fowkes, B. (Harmondsworth: Penguin Books in association with *New Left Review*, 1976, p.603.)

21. Marx, K. *Capital: a critique of political economy.* Volume One. Introduced by Mandel, E. Translated by Fowkes, B. (Harmondsworth: Penguin Books in association with *New Left Review*, 1976, p.1036.)

22. Marx, K. *Capital: a critique of political economy.* Volume One. Introduced by Mandel, E. Translated by Fowkes, B. (Harmondsworth: Penguin Books in association with *New Left Review*, 1976, p.615-616.)

23. Herodotus. *The histories.* Translated by de Selincourt, A. and revised by Burn, A.R. (London: Penguin Books, 1972.)

24. Warner, J. *From writing to computers* (London: Routledge, 1994.)

25. Marx, K. *Capital: a critique of political economy.* Volume One. Introduced by Mandel, E. Translated by Fowkes, B. (Harmondsworth: Penguin Books in association with *New Left Review*, 1976, p.617.)

26. Shannon, C.E. and Weaver, W. *The mathematical theory of communication.* (Urbana: University of Illinois Press, 1949, p.25.)

27. Ramsey, F.P. 'The foundations of mathematics.' In: Ramsey, F.P. *Foundations: essays in philosophy, logic, mathematics and economics.* Edited by Mellor, D.H. (London: Routledge and Kegan Paul, 1978.)

28. Cherry, C. *On human communication: a review, a survey and a criticism.* 3rd edition (Cambridge, MA: MIT Press, 1978.)

29. Wang, H. *From mathematics to philosophy.* (London: Routledge & Kegan Paul, 1974.)

30. Minsky, M. *Computation: finite and infinite machines.* (Englewood Cliffs, NJ: Prentice-Hall, 1967, p.104.)

31. Davis, M. 'Mathematical logic and the origin of modern computing.' In: Herken, R. (ed.) *The Universal Turing machine: a half-century survey.* (Oxford: Oxford University Press, 1988.)

32. Marx, K. *Capital: a critique of political economy.* Volume One. Introduced by Mandel, E. Translated by Fowkes, B. (Harmondsworth: Penguin Books in association with *New Left Review*, 1976.)

33. Marx, K. *Grundrisee: foundations of the critique of political economy (rough draft).* Translated with a foreword by Nicolaus, M. (London: Penguin Books in association with *New Left Review*, 1973, p.831.)

34. Martin, S.B. 'Information technology, employment, and the information sector: trends in information employment 1970-1995.' *Journal of the American Society for Information Science* 49(12), 1998, p.1058.

35. Benjamin, W. 'The work of art in the age of mechanical reproduction.' In: Benjamin, W. *Illuminations.* (New York: Schocken Books, 1969.)

### *Note*

[1] See Warner J. *From writing to computers.* (London: Routledge, 1994, pp.64-116.) for an exposition of a universal information machine (the universal Turing machine), for a discussion of automata theory and symbolic logic, and for a reiteration of the accepted view of automata theory and working computers as partly separate developments.

**Subscriptions**
*Aslib Proceedings* is published ten times per annum and supplied to corporate members of Aslib. For 1999 the annual subscription rates are £170/$280 (£136/$224 to members of Aslib). For subscriptions and back issues please contact: Portland Press Ltd, Commerce Way, Whitehall Industrial Estate, Colchester CO2 8HP, United Kingdom. Tel: +44 (0) 1206 796 351. Fax: +44 (0) 1206 799 331.

Aslib, The Association for Information Management, is a world class corporate membership organisation with over 2000 members in some 70 countries. Aslib actively promotes best practice in the management of information resources. It lobbies on all aspects of the management of, and legislation concerning, information at local, national and international levels. Aslib provides consultancy and information services, professional development training, conferences, specialist recruitment, Internet products, and publishes primary and secondary journals, conference proceedings, directories and monographs.

Further information can be obtained from: Aslib, The Association for Information Management, Staple Hall, Stone House Court, London EC3A 7PB, UK. Tel: +44 (0) 171 903 0000. Fax: +44 (0) 171 903 0011. Email: aslib@aslib.co.uk *http://www.aslib.co.uk/*

Periodicals Postage Paid at Rahway NJ and at additional mailing offices.
**POSTMASTER:** Send address changes to PROCEEDINGS, c/o Mercury International, 365 Blair Road, Avenel NJ 07001

©1999 Aslib and contributors ISSN 0001 253X

ISSN 0001 253X
Volume 51
Number 6
June 1999

25/2/2000
wei

# Aslib Proceedings

## New information perspectives

### A Siberian reality check on open source information
Athol Yates and Nicholas Zvegintzov

### Survival strategies in the Learning Age – hybrid staff and hybrid libraries
Penny Garrod

### Non-use of business libraries and information services: a study of the library and information managers' perception, experience and reaction to non-use
Laura Brick

### Constructing 'users' in library and information science
Heidi Julien

Aslib THE ASSOCIATION FOR INFORMATION MANAGEMENT

*Aslib Proceedings* is a long-standing professional journal of international repute, which has a large and wide subscriber base. The journal brings currency, authority and accessibility to the reporting of current research, issues and debates in the broad area of information work. Above all the journal wishes to provide research and comment in a form that is easily and quickly understood: a fresh, rigorous, but unfussy, writing style is what is aimed for. The needs of busy practitioners (and academics) are very much in mind. Articles are refereed to meet the need for accredited and authoritative information. To ensure that this does not result in an unacceptable loss of currency, articles are speedily refereed by the journal's large, eminent and multi-disciplinary Editorial Board (assembled for just such a purpose). In exceptional cases they may be referred to an external panel of consultants. The journal's monthly publishing pattern helps in maintaining the currency of reporting. In consequence the submission of research in progress articles is encouraged.

The boundaries of information work are rapidly changing and widening. *Aslib Proceedings* has a role to play in the repositioning of the information profession by encouraging submissions from information managers, providers, packagers and consumers of all kinds – and especially those from the related information fields of journalism, electronic publishing, communication and Internet studies. An interdisciplinary Editorial Board has been assembled to assist in the process.

Prospective authors should submit articles of 2,000-6,000 words in length. Articles should be accompanied by a 100-150 word informative abstract and the numerical form of referencing should be adopted. In style and tone, submissions should fall between that of the journal's sister publications: *Managing Information* and *Journal of Documentation*. Articles should be furnished by email or on disc to David Nicholas, Editor, Aslib Proceedings, Department of Information Science, City University, Northampton Square, London EC1V 0HB. Email: nicky@is.city.ac.uk. The Editor is happy to provide advice on an informal basis to potential authors.

**Editorial Board**
Dr David Nicholas, *Head of Department, Department of Information Science, City University;* John Akeroyd, *Director, Learning Resources, South Bank University;* Sarah Blair, *Head of Publications, Aslib;* Peter Chapman, *Information Manager, NewQuest North East;* Professor Peter Cole, *Head of Department, School of Journalism, University of Central Lancashire;* Dr Tom Dobrowolski, *Senior Lecturer, Institute of Information Science and the Book, University of Warsaw;* Harry East, *Senior Research Fellow, Department of Information Science, City University;* Mike Hyman, *Managing Director, Head Software International;* Robin Hunt, *Creative Director, AreHause;* Nat Lievesley, *Network Manager, Centre for Policy on Ageing;* Helen Martin, *Information Manager, Research and Information Unit, The Guardian*; Dr Ian Rowlands, *Lecturer, Department of Information Science, City University;* Richard Withey, *Director, New Media, News International;* Ralph Adam, *Freelance Journalist.*

**Editorial Team**
*Editor:* David Nicholas
*Deputy Editor:* Ian Rowlands
*Editorial Manager:* Kay Nicholas
*Editorial Assistants:* Peter Williams (City University), Colette Coles (City University)

# CONTENTS

175-186 A Siberian reality check on open source information
Athol Yates and Nicholas Zvegintzov

187-194 Survival strategies in the Learning Age – hybrid staff and hybrid libraries
Penny Garrod

195-205 Non-use of business libraries and information services: a study of the library and information managers' perception, experience and reaction to non-use
Laura Brick

206-209 Constructing 'users' in library and information science
Heidi Julien

# A Siberian reality check on open source information

*Athol Yates and Nicholas Zvegintzov*[i]
*Institution of Engineers, Australia National Office, Engineering House, 11 National Circuit, Barton ACT 2600, Australia*
*Email: athol_yates@eol.ieaust.org.au*

**Abstract** *Open Source Information (OSI) and Open Source Intelligence (OSINT) are attracting enormous interest from the business, military and political intelligence communities. OSI and OSINT offer the prospect of delivering valuable intelligence, from so-called open sources such as newspapers, experts, and online databases. While OSI and OSINT may offer considerable potential when employed to produce information on foreign countries, they are totally dependent on the researchers' and analysts' understanding of the target country's history, politics and society.*

*A research project on the BAM (Baikal-Amur Mainline) Railway in eastern Siberia provides a practical example of OSI. The project involved researching a book on the BAM (Baikal-Amur Mainline) Railway in eastern Siberia using OSI alone as it was impossible to visit the region. The 3,084km BAM traverses an almost unpopulated region which had been virtually closed to Westerners since it became a major gulag zone in the 1930s. Once the book's research was underway, it became clear that much of the English language published material was ill-informed and contradictory. After the collapse of the Soviet Union and the relaxation of travel restrictions, it was possible to validate the information gathered. This led to a number of surprising findings about the accuracy and comprehensiveness of the material collected outside Russia.*

## Introduction

Open Source Information (OSI) and Open Source Intelligence (OSINT) are the latest trends sweeping the business, military and political intelligence communities. Conferences are being held on them, large organisations are starting to employ them in a strategic fashion and academics are researching the techniques[ii]. OSI and OSINT offer the prospect of delivering valuable intelligence, from so-called open sources such as newspapers, experts, online databases, commercial imagery and technical publications, at a fraction of the cost of conventional intelligence techniques such as industrial espionage, spies and satellites.

While OSI and OSINT are not new concepts, they have been largely ignored for the last 40 years as traditional espionage tools have dominated the intelligence gathering process[iii]. OSI and OSINT were rediscovered in 1992[iv] and since then have increased their profile. The main exponents of them, Robert Steele and Mark Lowenthal, state that:

> 'OSINT results from the integration of legally and ethically available multi-lingual and multi-media sources (ie OSI), with the heretofore largely secret processes of national intelligence: requirements analysis, collection management, source validation, multi-source fusion, and compelling presentation'.[1]

A research project on the BAM (Baikal-Amur Mainline) Railway in eastern Siberia is a practical example of OSI. The project involved researching a book on the BAM (Baikal-Amur Mainline) Railway in eastern Siberia using OSI alone as it was impossible to visit the region. The 3,084km BAM traverses an almost unpopulated region which had been virtually closed to Westerners since it became a major gulag zone in the 1930s. Once the book's research was underway, it became clear that much of the English language published material was ill-informed and contradictory. While a few of the errors it contained may reflect highly on the creators of disinformation, most of it appeared to be due to sloppy research and reporting.

After the collapse of the Soviet Union and the relaxation of travel restrictions, it was possible to validate the information gathered. This led to a number of surprising findings about the accuracy and comprehensiveness of the material collected outside Russia. The analysis of the project is divided into three main components of OSI: discovery, discrimination and validation. From an OSI perspective, nine lessons can be drawn from the project. While the paper is based on experience in the Soviet Union and Russia, the lessons are equally relevant to other non-English speaking countries.

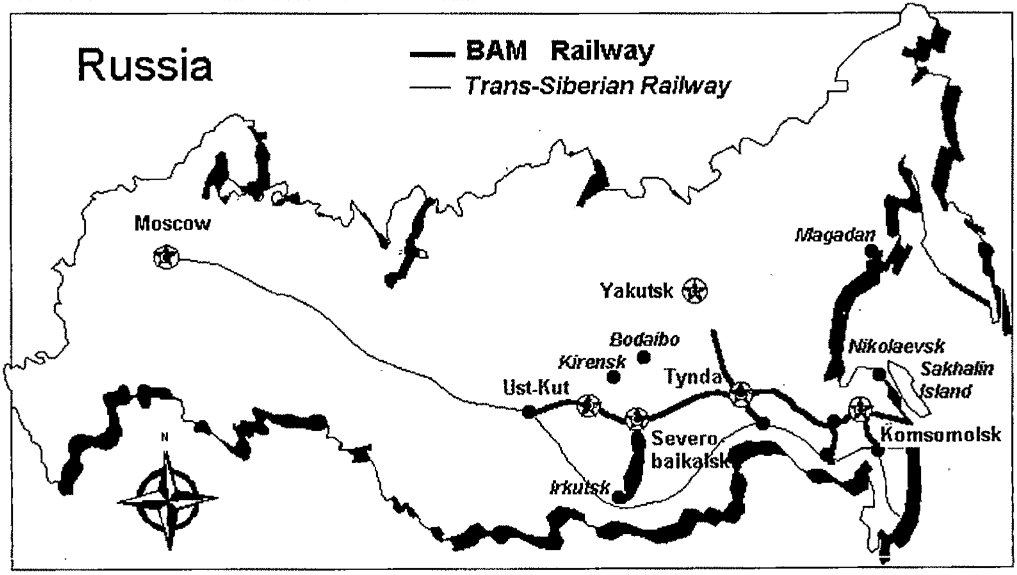
## Background

The research project resulted in the book, the *Siberian BAM Railway guide*[2]. While it could be categorised as a guidebook, in reality the book is a regional, historical and cultural study of a 3,084km long railway development zone in eastern Siberia. This is illustrated by the type of readers the book attracts. The vast majority of the book's readers are students of Russia rather than actual travellers. In fact, as less than 90 Westerners travel on the railway each year, travellers do not even make up a significant percentage of the book's buyers[v]. The book's authority is illustrated by the fact that it formed the basis for the entry on the BAM (Baikal-Amur Mainline) Railway in the *Modern encyclopedia of Russian, Soviet & Eurasian history (supplement)*[3] and a US company, which produces business, political and social databases on Russia, have proposed making the guide into an electronic encyclopedia on the region.

The BAM traverses eastern Siberia and runs roughly parallel with, and north of, the Trans-Siberian railway. The BAM provides Russia with a second railway link to the Pacific Ocean. The backbone of the BAM is the 3,084km railway line starting near the Lena River port of Ust-Kut in the west and ending near Komsomolsk-na-Amure in the east. The BAM includes three branch lines: AYAM (Tynda – Berkakit – Aldan and eventually Yakutsk), Little BAM (Tynda – Bamovskaya) and Urgal – Izvestkovaya.

The construction of the BAM was one of the great 20th century construction feats as it traverses 11 large rivers, seven mountain ranges and more than 1,000km of permafrost. It has eight tunnels, 142 bridges more than 100m long and more than 200 stations and passing loops. Construction of the railway

Figure 1: The BAM Railway and its location in Russia



started in the 1930s using prison labour from Stalin's gulags, but was discontinued when the camp doors opened following the dictator's death in 1953. Work resumed under Brezhnev in 1974 and the railway was officially opened in 1984. However, it was another six years before the railway was upgraded to a standard which allowed regular services to run along its entire length. Even today, the BAM is not complete as one tunnel still has to be finished.

The senior author became interested in the BAM in the late 1980s after reading conflicting reports about the existence of the railway and seeing in atlases that the BAM followed several different routes. It was known that the Soviets were good at disinformation but how was it possible in the 1980s that there was no consensus on the existence and route of a 3,000km railway? This question led to researching the BAM.

## Discovery

The first step in answering the question of the existence and route of the BAM was to examine general publications. At one end of the spectrum was the claim that the railway was completed in the 1940s, 40 years before it actually was. An example of this was the article in *The New York Times* on 11 August 1950.

The mistaken belief that the line was completed can be seen in maps of that era. One such map is found in *Forced labor in Soviet Russia*[5] (1947).

**The New York Times**

**August 11, 1950**

**Soviet Completes Far East Rail Link**

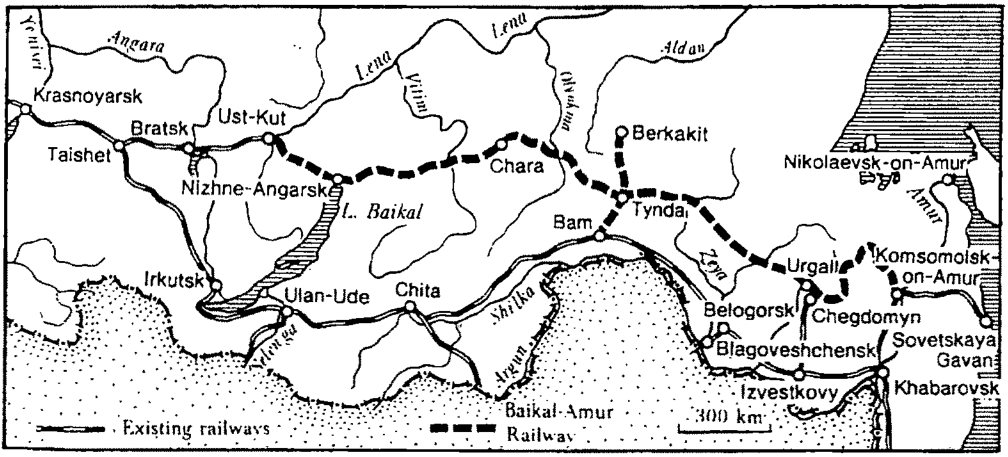
Soviet Russia has completed another strategic railroad to the Soviet Far East, paralleling the route of the double-tracked Trans-Siberian Railroad, Prof. Andre Karpinsky, exiled Russian geologist, reported here yesterday.

Professor Karpinsky said that the Baikal-Sovietskaya Gavan Railroad had been completed before he left Russia in 1943 and that students of his had made the geological surveys of the road.

Later books acknowledged that the BAM was not complete but drew proposed routes which were incorrect. For example, in *An atlas of Soviet affairs*[6] (1965), the dashed lined shows a 'tentatively proposed' line running north from Tynda to Yakutsk and then east to Magadan on the coast. In reality, the Tynda-Yakutsk-Magadan route was never a serious contender. To compound the error, the map does not show the Tynda-Komsomolsk route which became the future route.

It is interesting to note that both the map in *Forced labor in the Soviet Union* and *An atlas of Soviet affairs* showed the western end of the line passing through the gold mining

*Figure 2: A close up of the 3,084km BAM Railway route[4].*



centre of Bodaibo, which is over 300km to the north of the line's actual route. This significant error was often repeated. For example, in *The first guidebook to prisons and concentration camps of the Soviet Union* (1980) it states that Bodaibo is 'accessible by train from Taishet'[7].

The most recent repetition of this erroneous information appears in the 1989 *Trans-Siberian rail guide*[8]. This showed that the BAM will pass through Bodaibo. It also illustrated that the BAM had already been built to Kirensk which was a significant error as the railway neither had been built to that town nor ever would be. These two errors indicate that either the travel writer was confused or never travelled on the line personally.

Other recent maps have been more accurate, such as the *USSR: a travel survival kit* (1991)[9], which correctly showed the basic BAM route but also contained two significant errors. Firstly the map showed the BAM as only partially complete at a time when it had been finished for six years (with the exception of three tunnels) and secondly, it omitted the three north-south lines which connect the BAM with the Trans-Siberian. At the other end of the spectrum of claims about the existence of the BAM were maps which completely

*Figure 4: A map from* An atlas of Soviet affairs *which incorrectly shows the proposed route of the BAM.*

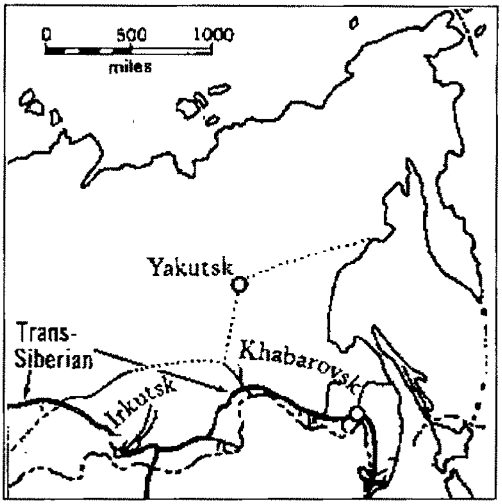
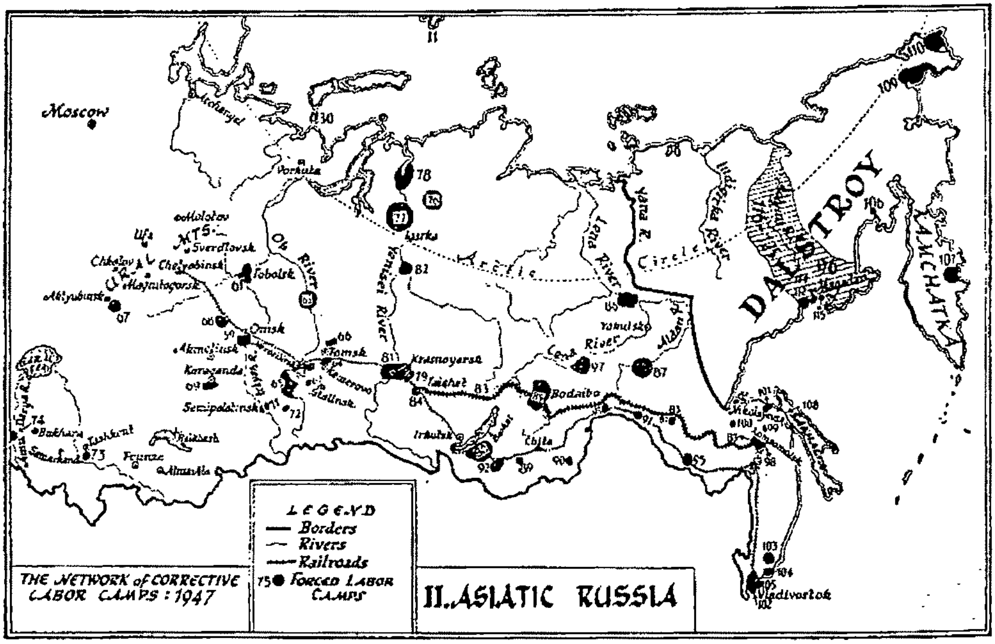


*Figure 3: A map from* Forced labor in Soviet Russia *which incorrectly shows the BAM was completed.*



omitted the BAM, such as *The new jacaranda atlas* (1989)[10].

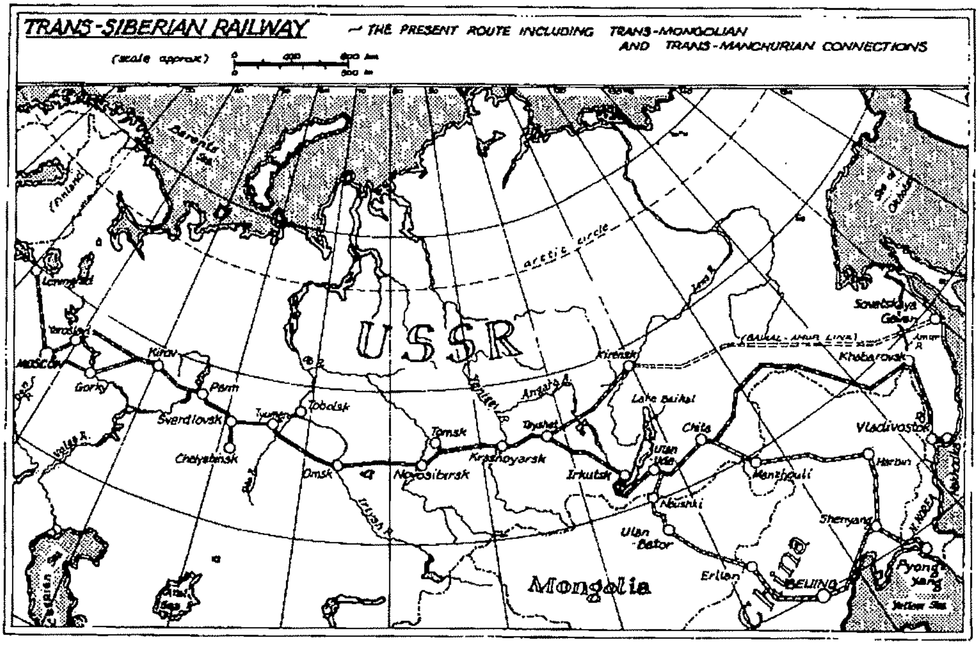
Once it became obvious that most general English language publications were inadequate or contradictory, the authors started searching for high quality, primary source material. It soon became apparent that there was a very limited range of English language translations of primary sources. The only useful ones which had occasional articles on the BAM were the *Current Digest of Soviet Press* and *Soviet Geology*.

A search of English language academic writing on the BAM identified only one subject matter expert on the BAM; the academic Victor Mote[vi]. Victor Mote's analyses, which regularly appeared in *Soviet Geology* revealed a long-term interest in the region. Without his work, collecting background information on the BAM would have been extremely difficult. This led to what, for the purposes of this paper, is called OSI Observation 1 which is that subject matter experts probably exist on all subjects and they are essential in providing a reliable base of information on which to build analysis.

A search of Russian language material on the BAM revealed that it was extensively covered in many books and periodicals. Books devoted to the BAM including *Magistral Veka*[11] (The Great Baikal Railway) (1977) and *BAM Prodolzhaetsya*[12] (The BAM Continues) (1990). Periodicals with regular articles on the BAM included *Gudok* (The Whistle), *Izvestiya* (News), *Stroitelnaya Gazeta* (Construction Journal), *Zheleznodorozhny Transport* (Railway Transport), *Ekonomicheskaya Gazeta* (Economic Journal), *Voprosy Ekonomiki* (Questions of Economics), *Ekonomicheskaya Geografiya SSSR* (Economic Geography of the USSR) and *Narodnoye Khozyaystvo* (National Economy).

The route of the BAM could be found in many Soviet publications from the mid 1970s onwards. For example, the 1977 edition of the *Bolshaya Sovetskaya Entsiklopediya* (Large Soviet Encyclopedia) shows the detailed route.

*Figure 5: A 1989 map from the* Trans-Siberian rail guide *which contains two significant errors. Firstly, the 'completed' line to Kirensk never existed and secondly, it shows that the proposed BAM line passes near Bodaibo, which is 300km to the north of the line's actual route.*



This discovery led to OSI Observation 2 which is that specialised primary information on a remote and closed area of the Soviet Union did exist, but the vast majority of it is not in English.

During the early 1990s, on-line sources in Russian were of no use as none of the primary sources were in electronic format. By 1998, there are several relevant on-line Russian language sources. However, compared to printed material, they do not represent a significant source of historical or current information on the BAM. This finding led to OSI Observation 3 which is that highly specialised on-line sources in Russian were very limited and are still not significant today.

Considering the plethora of accurate Russian language primary material, there was no excuse for English language books to publish inaccurate information on the BAM by the 1980s. The very fact that they did suggests that the authors never undertook serious research, but instead just reprinted inaccurate material from other English language authors. This lead to OSI Observation 4 which is that errors are often repeated in open source information on Russia as the foreign language discourages writers to use primary sources to verify previously printed English language reports.

## Discrimination

Discriminating between various sources of information was vital, particularly considering that it was known that some of the English language material was inaccurate. Sometimes it was easy to identify material which had a high probability of being inaccurate and therefore should be discarded. For example, the following quote contains the unusually frank admission that the source of the information was a second-hand rumour, and warns readers of the source unreliability.

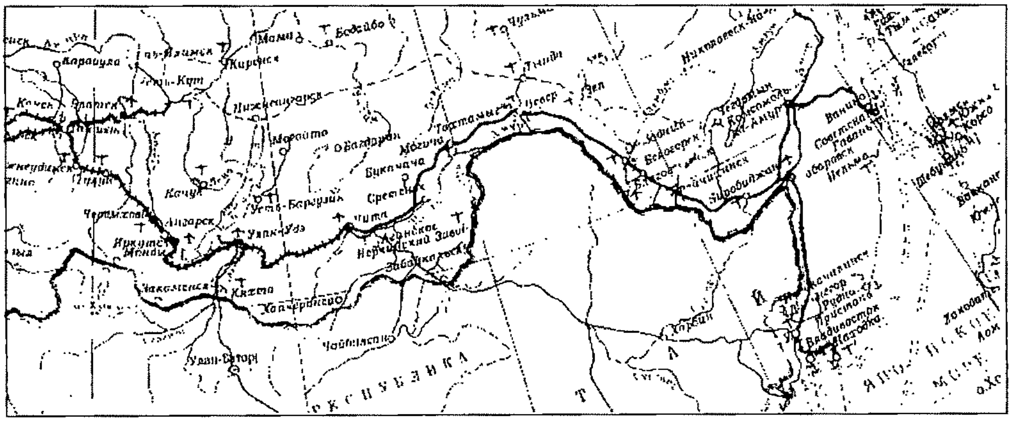
> 'An American traveller in Siberia brought back the story in 1941 of the building of another railway line in the Soviet Far East. He did not see it himself, only heard about it from a young Communist. This line was to have its southern terminus at Komsomolsk and from there run along the Okhotsk Sea, eventually reaching Kamchatka peninsula and the Bering Sea.'[13]

Incidentally, discarding this information has been justified as the authors have not discovered any serious Soviet-era plans for this route.

Several techniques to discriminate between good and bad sources, current and outdated sources, and relevant and irrelevant sources were used. Firstly, if the source of the material was given, the material was worth further consideration. Source attribution made it possible to drill down to the original source so that it could be evaluated.

Comparing the quoted material with the original material revealed that quoting authors

*Figure 6: A map which appeared in the* Bolshaya Sovetskaya Entsiklopediya (1977) *correctly showing the proposed BAM Railway route.*



occasionally took considerable poetic licence. An example of this was an article on the BAM published in *Moscow Times* on 18 July 1998. It stated that:

> 'The BAM zone began attracting investors in the mid-1800s when explorers considered and then gave up construction, concluding that it was too complicated and expensive. Then a French entrepreneur, Loik de Lobel, representing US rail companies, proposed to finance the construction of a Siberian-Alaskan railway in exchange for a 90-year-lease of all land within 25 kilometers of the line, according to Yates.'

It would appear from this quote that the person being referred to (i.e. the senior author) considers that there was a link between the BAM Zone and the proposal by Loik de Lobel. However, this was not the case as the journalist took some licence with the statements on which her quote is based. Her quote was based on page nine of the *Siberian BAM Railway guide*. The statements were discussing the history of railways in Siberia and were not referring to the BAM Zone directly. The statements were:

> 'One of the most famous development proposals [for a Siberian Railway] was from a French entrepreneur, Loik de Lobel, acting on behalf of US rail companies in 1904. He proposed a Siberian-Alaskan railway starting near north Baikal, through Yakutsk and then across the Bering Straits to Alaska. The construction would be privately funded on the condition that the bankers had a 90 year lease on all land within 25 km of the railway.'

Had Loik de Lobel's proposed railway gone ahead, it would not have passed well to the north of the BAM Zone. This led to OSI Observation 5 which is that it is necessary to check the original source before accepting quoted material.

Secondly, sources needed to be weighed against one another with more reliability attributed to newer and specialist publications and less to older and generalist publications. This reflects the fact that until the 1960s, most information on the region came from former gulag prisoners and prisoners of wars. Reports from these people did not provide sufficient information to build a region-wide view of BAM developments. After the 1960s, spy satellites could provide accurate information but this appears to have not been released to researchers as illustrated by the incorrect maps in publications such as *An atlas of Soviet affairs*[14]. By the late 1970s, the information on the BAM was very reliable as films, articles and first hand observations were readily available.

Thirdly, unsourced material was noted and accepted if it was collaborated by independent material. For example, a book on Soviet prison camps claimed there was a mining gulag camp near Severobaikalsk at Lake Baikal's northern end. Despite the observation that prison camp books normally contained considerable errors, the mining claim was cross referenced against books on the natural resources of Siberia. These stated that mica existed at both the northern and southern ends of Lake Baikal. It was also known that mica was mined in the south. Therefore it was reasonable to believe that a mica mine existed near Severobaikalsk and that its miners were probably gulag inmates. This information was later validated first hand when the author was taken to the remains of Akikan Gulag camp. This point led to OSI Observation 6 which is that unsourced information is useful if it can be independently collaborated.

Analysing seemingly unrelated material is a useful way to make insightful deductions. For example, in 1941 *Pravda*, reported that chickens were airfreighted from Irkutsk to Bodaibo. At the time, the BAM was thought to go through Bodaibo. This news revealed that the line to Bodaibo was either not completed or was not being completed. This assertion was supported by a report in *Pravda* on 2 November 1942 which stated that the people of Yakutsk donated presents for the soldiers at the front and these were despatched by train from Irkutsk. This implied that they were flown from Yakutsk, over Bodaibo to Irkutsk. This led to OSI Observation 7 which is that seemingly unrelated information can be useful in deducing and collaborating other information.

The discovery and discrimination phases are extremely time consuming. As an illustration of the large number of sources examined during the two phases, over 420 publication request slips were lodged with the National Library of Australia alone.

### Validation and local discovery

In the early 1990s, Russia relaxed travel restrictions on foreigners which gave the senior author the opportunity to travel to the BAM to validate the research. The research notes were ringbound into 200 A4 pages. The information was organised into rough chapters which were carried over into the final guide. These chapters were:

- the history of the BAM;
- life in the BAM Zone;
- rail operations;
- planning a trip;
- a summary of each village, town and city in the BAM Zone.

On the road, corrections, new information and questions were added to the notes. As can be imagined, the senior author was paranoid about losing this much-scribbled book. Three times this nearly occurred. The first time was when the author was interviewed by the KGB in the engine cab of a locomotive. This followed an attempt to take photographs from the cab as the train crossed over a bridge. The military guard on the bridge saw this breach of law and contacted security at the next station at which the train was to stop. The second incident was when the author's photographer was taking photographs of Khabarovsk station and was approached by a heavily armed station guard with an attack dog. The third incident was when a policeman saw an accurate hand drawn map of Taksimo township in the ringbound book and thought it was a 'state secret'. A combination of pleading and showing an impressively stamped letter of support from Moscow diffused each situation.

Validating physical data was relatively easy as the reported spur line, monument or hotel was either there or not. However when it came to validating history or non-observable things, it was considerably more difficult as it relied mostly on people's opinions and on people providing leads for collaborating evidence.

From the experience of questioning locals, the can be grouped into four types. The first type were those who repeated and embellished second-hand information without actually knowing the facts. For example, one local stated that to get to the Marble Canyon uranium mining gulag camp from Novaya Chara station would take about one day. In reality the round trip takes eight solid days of trekking, including crossing a marsh, a glacier and several rivers.

The second type were people who knew but either pretended not to know or refused to say. There were several reasons for this. These included protecting information which could put the country in a bad light (this means great slabs of history are ignored), regurgitating the Party approved line (despite the fact that the there was now no Party) and self interest. An example of a person who refused to state the facts was the guide to the BAM Museum in Novy Urgal. When asked to talk about prisoners who built the Dusse-Alinski tunnel in the 1940s and 1950s, she stated that only soldiers built the tunnel and it was only built in the 1970s. However, this was not the truth. The tunnel was actually completed by gulag prisoners in 1950 but became blocked after it was abandoned in 1959 when it became obvious that the railway line to the tunnel was not going to be built. The soldiers stayed in the former gulag camp buildings and found frozen corpses and other remains of the former tunnellers. Soldiers started sending the gruesome findings home and the most famous of these was a skull turned into a candlestick holder. This scandal was reported in the Soviet press in the mid 1970s and resulted in a military commission investigating the situation. This commission resulted in the destruction of all gulag-era evidence and the burying of the story. So despite the well reported story, the guide still perpetuated the Soviet-era version of history.

The third type were people who knew surprisingly little about their immediate region. There were several reasons for this including a societal bias against historical curiosity, extremely high labour turnover, limited ability to travel, and obsessive secrecy. An example of this was the refusal by Russian authorities to release a strategic management plan for the Lake Baikal region which was developed during the 1990 Expertise Conference held in Severobaikalsk. Conference participants included UNESCO, the Californian Earth Institute, Soviet agencies and local representatives. In this particular case, locals got around the problem by obtaining an English language copy of the report, which was freely available in the US, translating it and getting it published in the local Severobaikalsk newspaper.

The fourth type were people who had first hand knowledge. For example, one local drew a map of the very small, isolated town of Taksimo which would have been next to impossible to obtain without her help.

As it is very difficult to determine which type a person is when they are providing (or not providing) information, the only way to guarantee information accuracy is to confirm it first hand. This is OSI Observation 8.

Other local sources of information included museums, archives, photograph albums, libraries and local newspapers. However there were significant limitations on the usefulness of these sources. Despite the years of *glasnost*, most museums had only Soviet-era propaganda with a little information about the political repression of Stalin. Access to archives was very difficult, most of all because there was none in the BAM Zone itself. For example, the Komsomolsk-na-Amure archives were held in Magadan which is over 1,400km to the north of Komsomolsk-na-Amure. Photograph albums of natural and man-made wonders were useful on the railways and the region. Pictures of monuments led to interesting historical sites, aerial photographs of towns allowed street maps to be drawn and illustrations of permafrost protection led to the discovery of fascinating engineering techniques.

Library books were of limited use as all libraries had the same homogenous Soviet-era holdings. However libraries did have two useful sources of information. Firstly, some libraries held specialised maps, such as fishing or hiking maps. A few finds were very lucky, such as a topographical map of the Kodar Mountain Range[15] designed for serious alpinists, hikers and cross country skiers. It covered the area of the abandoned Marble Canyon uranium mine gulag camp and although it was not referred to directly, the gulag's site was marked on the map as 'Geologists' settlement (abandoned)'. Secondly, libraries held limited circulation local newspapers. These newspapers, often as small as a one tabloid page and produced only once a month, were published in Severobaikalsk, Taksimo, Chara, Tynda and Novy Urgal, Vanino/Sovetskaya Gavan and Komsomolsk-na-Amure. In addition, there was one local railway newspaper, not surprisingly called *BAM*. None of these publications were held in public libraries outside Russia at that time, according to the Australian Bibliographic Network. This led to OSI Observation 9 which is that much of the specialised primary source material of Russia is not available outside that country. As the BAM Zone libraries did not have photocopiers, researching was very time consuming.

One of the greatest finds during the local research phase was a secret Soviet report[16] written in 1945. There were 200 numbered copies printed of the book and it showed the exact route of the BAM. The route in the report was virtually identical to today's route. There were two main implications of this report. Firstly, as the route did not go northwards to Bodaibo, Western reports of the Bodaibo route (up until 1989) were not based on reliable Soviet sources. This meant that they were either guesswork or based on Soviet disinformation. Seeing that it was impossible to hide a railway with the advent of spy satellites in the 1960s and that the railway was built to Severobaikalsk (which was the southern route) in about 1981, the errors could not have been due to Soviet disinformation. Therefore the route error was due to sloppy research.

The second implication was confirmation of Soviet press reports about the failure to undertake proper research before digging a 15.7km tunnel through the North Muya Mountain Range. The 1945 map showed the tunnel at the same site at which work started three decades later. Considering the advances in geology over 30 years, it is surprising that this was still the optimal route. This observation supported the 1980's claim made in the Soviet press that the decision to build the North Muya tunnel was fundamentally wrong. According to Vladimir Ignatovich, chief geologist for the Buryat Geological Production Association (Buryatgeologia) which surveyed the original BAM line:

> 'Back when the route was being surveyed, we warned the designers at Novosibirsk's Siberian State Transport Design Institute and the Leningrad State Subway Design Institute about the highly complex conditions in the BAM's Buryatiya sector. We felt it would be better to bypass it from the south with minimal excavation work. However the shortest part was chosen, a 15km tunnel.'

Therefore, it would appear that the site for the tunnel was not based on surveys of the 1960s and 1970s but rather on the 1945 master plan.

The North Muya Mountain Range tunnel is also an example of local research neither validating nor disproving a BAM story frequently published in the West. The story is based on an accident in September 1979 when a drilling team hit an unexpected fault line which held 12,000 cubic metres of water, sand and rocks. The water surged into the drilling gallery in just a matter of seconds, drowning a number of miners. Details of the disaster were published in *Pravda* on 28 February 1983 but the report contained no mention of deaths. A much quoted report put the toll of dead and injured at several hundred[17]. The author made several different attempts to find the true figure but these were unsuccessful. Firstly, the co-author, a former tunneller who worked elsewhere on the BAM at the time, contacted his former colleagues but none knew the death toll. Secondly, contact was made with one of the railway construction managers of the tunnel at the time but he refused to put a figure on it. Thirdly, a search for memorials to the tunnellers was fruitless.

Local research also revealed several local stories which were not (unsurprisingly) mentioned in Western publications. These included a UFO club at Seti and Yeti sightings in Vanino. However the most significant local discovery, which was not mentioned in any literature during the discovery phase, was a planned 400km northern extension and a 7km undersea tunnel to Sakhalin Island. In Komsomolsk-na-Amure, the historian Marina Kuzminovna, told the author about a 1950 secret order by the Council of Ministers to build a 400km railway from near Komsomolsk-na-Amure to the Cape Lazarev on the Tatar Straits. The only other information she provided was that 122km of the line was laid before Stalin died and the project was abandoned. In Nikolaevsk-na-Amure, the author met the local museum head who had started visiting and collecting information about the gulag complex in Cape Lazarev in the early 1990s. He had seen the remains of the gulag camp and a vertical tunnel shaft which was the only tunnel works done before the project was abandoned. As an illustration of how compartmentalised information was in Russia, at the same time the author was travelling, an article appeared in the Russian railwayman's newspaper *Gudok* asking readers if they had any information on the alleged Tatar Strait tunnel. This lack of knowledge is quite surprising considering that the newspaper is 90 years old and has the support of the two million strong State railways behind it.

One of the hardest aspects of developing a guidebook is to produce town maps. Soviet maps are intentionally inaccurate, a fact which dates back to a decision in the 1930s when the government decided that accurate maps could be used by an invading enemy. This decision still stood, despite the advent of spy satellites, until 1989 when the Main Administration for Geography and Cartography (GUGK) produced the first accurate map of Moscow in over five decades. To produce maps of villages, towns and cities, a number of approaches were used. Locals were asked to draw maps of their village, blueprints of a city were obtained via connections with the son of the city's mayor, and aerial maps of cities from photograph albums were manipulated. Only once did the author resort to using a compass and pacing but it soon became obvious from the stares of the onlookers that this could lead to being arrested for spying. It must be remembered that taking photographs of bridges, railway installations and other 'sensitive' installations was and is still forbidden in Russia.

The one category of information collected in the West which was completely wrong was information on services for travellers. Just prior to the research starting for the second edition of the guidebook in 1998, the update author was advised by numerous people that the vast majority of services for travellers, such as hotel accommodation, ferries and restaurants, were no longer available. Local research showed that the services were still available, although some of them were sporadic. There were several reasons for the inaccurate information. Firstly, the people providing the advice did not have first hand information. Secondly, communication in the BAM Zone is quite difficult, as there are two different telephone systems among other reasons, and if no answer was received, it was assumed that the service was no longer available. Thirdly, reliable information on services is only available from the location which organises the service. For example, ferry timetables can only be trusted at the point of the ferry's departure as information at the destination and points along the way may be old, invented or completely unavailable.

The net result of the validation phase was that about 30% of secondary source and about 90% of primary source information turned out to be accurate. In addition, the book grew by about 20% as a result of new information discovered while in Siberia.

## Conclusion

As a case study of OSI, the researching and writing of the *Siberian BAM Railway guide* is an interesting example. The research was done in an era when on the ground researching was not possible but by the time it was finished, the area was opened and the collected data could be verified.

With hindsight and a lot of money, it would have been possible to produce nearly all of the book without visiting the region. This would have been done by purchasing satellite imagery, interviewing foreign technical experts who supplied mining goods to the region, and analysing Russian publications. However on the ground research would have been needed to complete the maps, obtain information on services for travellers, and gain access to locally produced newspapers where not held in the West.

The lessons learned from an OSI perspective in this project include that:

- subject matter experts probably exist on all subjects and they are essential in providing a reliable base of information on which to build analysis;
- specialised primary information on a remote and closed area of the Soviet Union did exist but the vast majority of it is not in English;
- highly specialised on-line sources in Russian were very limited and are still not significant today;
- errors are often repeated in open source information on Russia as the foreign language discourages writers to use primary sources to verify previously printed English language reports;
- always check the original source before accepting quoted material;
- unsourced information is useful if it can be independently collaborated;
- seemingly unrelated information can be useful in deducing and collaborating other information;
- it is necessary to obtain first hand confirmation of verbal information supplied by locals; and
- much of the specialised primary source material of Russia is not available outside that country.

## References

1. Steele, R. and Lowenthal, M. *Open source intelligence: executive overview.* (Fairfax, VA: OSS Inc., 1998, p.5.) [According to Steele and Lowenthal, OSINT is based on premises including that the private sector now controls roughly 90% of the pertinent information for creating both open and secret intelligence; and that within the private sector there are niche experts on any given topic and any given source whose expertise cannot be matched within any given intelligence community.]

2. Yates, A. *Siberian BAM Railway guide: the second Trans-Siberian railway.* (London: Trailblazer

Publications, 1995.) [The second edition, edited by Nicholas Zvegintzov, is due out in 1999.]

3. 'BAM Railway' In *Modern encyclopedia of Russian, Soviet and Eurasian history (supplement).* (USA: Academic International Press, 1998.)

4. Zhuravlev, V. *A second Trans-Siberian.* (Moscow: Progress Publishers, 1980, pp.11.)

5. Dallin, D. and Nicolaevsky, B. 1947, *Forced labor in Soviet Russia.* (Yale University Press, 1947, pp.65.)

6. Kingsbury, R. and Taaffe, R. *An atlas of Soviet affairs.* (Frederick Praeger, 1965, pp.119.)

7. *The first guidebook to prisons and concentration camps of the Soviet Union.* (Batam Books, 1980, pp.156.)

8. Strauss, R. *Trans-Siberian rail guide.* (UK: Bradt Publications, 1989, pp.viii.)

9. *USSR: a travel survival kit.* (Melbourne: Lonely Planet, 1991.)

10. *The new jacaranda atlas.* 3rd ed. (Australia: Jacaranda Press, 1989).

11. Malasenko, *Magistral Veka.* (Moscow, 1977.)

12. Koposov, G. and Koreshkov, V. *BAM prodolzhaetsya.* 1990.

13. Lengyel, E. *Siberia.* (New York: Random House, 1943, pp.311.)

14. Kingsbury, R. and Taaffe, R. *An atlas of Soviet affairs.* (Frederick Praeger, 1965, pp.119.)

15. Khrebet Kodar-Zabaikalskie Alpy, 1990, Turistskaya Karta, Moskva.

16. NKVD, 1945, *SSSR Upravlenie Po Uzyskaniyam i Proektirovaniyu Baikalo-Amurskoi Zheleznodorozhnoi Magistrali*, BAMPROEKT, Baikalo-Amurskaya Zheleznodorozhnaya Magistral. Komsomolsk-na-Amure. (People's Commissariat for Internal Affairs, 1945, USSR Administration for the Investigation and Planning of the Baikal-Amur Mainline, BAM Planning Organisation, Komsomolsk-na-Amure.)

17. Whiting, A. *Siberian development and East Asia: threat or promise.* (Stanford University press, 1981, pp.61.)

## Notes

[i] Athol Yates, BEng, GradDip Soviet Studies, is the author and Nicholas Zvegintzov is the update editor of the *Siberian BAM Railway guide: the second Trans-Siberian Railway.* 2nd ed.(London: Trailblazer Publications, 1999.) Athol Yates is a Policy Analyst, Institution of Engineers, Australia and Nicholas Zvegintzov is the President of Software Maintenance News Inc. (USA).

[ii] An example of this is the Optimising Open Source Information Conference held on 7-8 October 1998, Canberra, Australia which included papers on OSI and OSINT by intelligence professionals including Major General Bill Crews, Director of the Defence Intelligence Organisation (DIO), Jason Brown, Australian Security Intelligence Organisations (ASIO), and Dr V. Lakshmi Narasimhan, Defence Science and Technology Organisation (DSTO).

[iii] Governments in particular are re-evaluating this traditional approach to intelligence as it results in the production of classified data which is often very expensive, available on only a narrow range of topics, is out of data by the time it is read by the end-user, and its reliability cannot be assessed easily by the end-user as its sources are withheld.

[iv] Robert Steele and Mark Lowenthal founded the OSS Inc. and OSS Academy in 1992 and state, somewhat self-servingly, this was the year when the 'OSINT Revolution Begins'. Robert Steele probably has the highest world-wide profile for promoting OSI. He does acknowledge that OSI is not new as legal travellers gathered intelligence even before the Roman Empire, however it has been neglected since the Second World War.

[v] During the first nine months of 1998, only 40 Western tourists had stayed in Severobaikalsk (the most popular city for tourists on the BAM) for more than one day. Personal correspondence with Rashit Yahin, head of BAMTours, Severobaikalsk's only tourist company, 28 September 1998.

[vi] Victor Mote, University of Houston, wrote the first detailed book on the BAM. The book, *Gateway to Siberian resources (The BAM).* (Scripta Publishing Co.) was written with Theodore Shabad and published in 1977. It is interesting to note that he only travelled once on the BAM. That was in March and April 1985, when he rode the BAM for 52 km from Ulkan to Kunerma. During this time he was accompanied by representatives of Novosti Press and the KGB. Despite this lack of on-the-ground experience, his connections with Soviet experts and other sources gave him reliable insights. Since 1991, he has worked on a broader Siberian-Russian Far East area. He has also written *An industrial atlas of the Soviet successor states.* (Industrial Information Resources, 1994) and *Siberia: worlds apart.* (Westview, 1998).

# Survival strategies in the Learning Age – hybrid staff and hybrid libraries

*Penny Garrod*
*EIS Office, The Library, University of Plymouth, Plymouth PL4 8AA*
*Email: pgarrod@plymouth.ac.uk*

**Abstract** *In the Origin of Species, Charles Darwin sets out his observations on the evolution of species, and survival through adaptation to changing environments. Darwin's theories are applied to the current learning environment, which has undergone radical changes. It is argued that academic libraries and library staff need to adapt swiftly to the new learning environment, now that education is high on the political agenda. Institutions of higher education face stiff competition from a range of non-traditional education providers, and information professionals also risk being usurped by other professions eager to become involved with the Internet. The hybrid information professional and the hybrid library are evaluated as survival strategies for the new learning environment. The history of the hybrid professional is traced through the literature, and initial findings from the hybrid library projects (Phase 3 of the Electronic Library Programme (eLib)) are evaluated.*

## Introduction

> 'The process of hybridization is important biologically because it increases the genetic variety (number of different gene combinations) within a species, which is necessary for evolution to occur. If climatic or habital conditions change, individuals with certain combinations may be eliminated, but others with different combinations will survive. In this way, the appearance or behaviour of a species gradually may be altered. Such natural hybridization which is widespread among certain species makes the identification and enumeration of species very difficult.'[1]

If Charles Darwin could board the Tardis or the Starship Enterprise, and travel or warp from mid-Victorian industrialised Britain to the digitised virtual world of the Information Age, he might be amused to observe how his ideas, articulated in *Origin of Species by Means of Natural Selection*[2], are being applied to a domain far removed from the biological sphere – the libraries of academe. The notion of the 'hybrid' library professional was once ruled out as a viable possibility, but is gradually gaining in acceptance as managers of academic libraries realise the networked information environment is here to stay, and a new breed of information professional is needed to support it. The hybrid library, as a model for current and future service delivery, is also under investigation and development under Phase 3 of the Electronic Library (eLib) Programme[3].

It could be argued that the term hybrid is widely used in a variety of contexts, remote from natural evolution and Darwinian ideology. 'Hybrid' simply signifies that which is a mixture of other things – usually two other things. When used by the Library and Information Services (LIS) community it therefore reflects an ability to operate in both the electronic and the print environment. Or does it? On one level it certainly does, but if we examine the national and global environment in which the modern academic library functions, we find an environment which has been evolving for some time, and which is likely to change even more dramatically in the future. The hybrid library and the hybrid information professional could be interpreted as survival strategies in response to a changing environment.

The political and cultural environment in which higher education now operates is a much changed one. It is articulated through sound bites and slogans: 'Information Superhighway'; 'The Learning Age'; 'University for Industry' and 'Education, Education, Education'. The trend away from ivory towered centres of excellence for an elite minority has accelerated as programmes of mass education become the norm, and traditional 'chalk and talk' methods of teaching are being questioned as technology opens up a raft of opportunities for delivering mass education. This environment anticipates citizens who are lifelong learners; 'information haves' who possess core transferable skills which are constantly upgraded to keep pace with the vagaries and demands of the global marketplace. Learning is on the political agenda, and Information and Communications Technology (ICT) is the agent of change, which makes new methods of learning possible.

## The new learning environment

The new learning environment looks set to become a volatile marketplace. Higher education institutions will face stiff competition from a host of non-traditional education providers in the race to provide UK citizenry with the vocational courses and core skills required by society, especially in the context of the University for Industry. Distance, flexible, and lifelong learning are heavily reliant upon ICT, as time and place of study no longer represent barriers to learning. However, these methods of learning require a high level of support in the form of training and advice, much of which will need to be offered via the telephone, email or via video conferencing. A critical success factor for education providers will therefore be the quality and level of learning support offered to distance, part-time and lifelong learners. In a world of diversification and co-operative ventures, academic libraries are having to rethink the way they deliver current services; they have to become less process oriented, and more service or user oriented. They are being asked to think in terms of resource sharing, marketing what they do well, value added services, and tailoring services to meet the needs of specific groups of learners. Library services therefore have to evolve to survive, but organisations and professional groups are notoriously slow to evolve – the process is gradual and incremental, especially in comparison to technological innovation. The survival of academic libraries in the Learning Age is therefore dependent not only on an ability to adapt and respond to the challenges ahead, but to do this quickly. The risk of being usurped by other players in the education field cannot be ignored; there are potentially many service providers, who are keen to harness new technology, and bring learning materials and information resources to the place of study, wherever that might be.

The professional literature abounds with warnings and strategies for the survival of the information professional. The eLib Programme itself was a response to the Follett Report of 1993, which looked at the factors affecting library staff, in particular the changing nature of the student population and 'the demands which certain types of students typically place on library staff'[4]. A year later the Fielden Report tackled staffing requirements in academic libraries[5]. Amongst the Report's recommendations on training and staff development was a call for the development and updating of IT skills and competencies. These reports already seem somewhat dated as they precede the explosive growth of the Internet, and the range of education and ICT initiatives launched by the new government.

The need for new programmes of education for information professionals is a popular and recurring theme. Van Brakel uses the Darwinian phrase 'survival struggle' to describe the scenario in which information professionals battle with other professions 'to obtain a visible and viable share of the information industry, or rather information marketplace'[6]. Van House and Sutton refer to the 'panda syndrome' when describing the fate that could befall a recalcitrant information profession if it fails to update its educational programmes[7]. The authors suggest a strategy for survival which includes retaining the generic material of librarianship as a knowledge base, whilst adding new areas of knowledge.

Darwin warned that specialisation bears an inherent risk. If the environment changes, those who have gambled on specialisation may lose out, whilst those who have retained a generalised form and remained adaptable can adjust to the new situation and survive. However, some librarians argue that information is information, whatever the format. Librarianship is one of the oldest professions, and has no need to prove professional competence in something it has done for centuries, i.e. organising and retrieving information. The response to this argument might be: 'its information, Jim, but not as we know it'. Librarians may have 'been there first', and do not need instruction in sucking eggs, but this attitude fails to acknowledge the fact that the whole environment has changed. Librarians do *not* have years of proven expertise in navigating the Internet or in evaluating massive resources on a global scale. The medium is still too new, too vast and growing by the minute. The web is interactive, end-user oriented, dynamic and constantly growing. As such it is regarded as 'sexy' and 'cool', and a world apart from libraries with their set procedures and rules. It is therefore attracting those whose mindset, aptitudes, and skills are different from those of the traditional librarian. The people who are interested in exploiting the Internet tend to be business oriented and proactive; they are unlikely to carry around years of professional baggage, or suffer the stasis of bureaucratic organisations, where creativity and rapid response to change are stifled. The traditional library and the traditional librarian are likely to have difficulty keeping up with these more aggressive competitors.

Libraries will have to update the services they offer, the knowledge and skills set of their staff, and their image, if they are to survive both the changing information environment, and the changing learning environment. Moreover, they will have to do this swiftly. The 'hybrid' member of staff, and the 'hybrid library', represent a move in the right direction and an acknowledgement of the need for change.

## The evolution of the hybrid member of staff

The notion of the hybrid member of staff has been a topic of debate for several years. The need for library staff to have IT skills came to prominence in the late 1980s, when several pioneering higher education institutions were in the throes of converging thir library and computing services[8]. Service convergence was seen to be a more effective way of delivering a seamless information service in a distributed, end-user, IT environment. Convergence gave rise to the need for staff who were able to support users of both the networked learning environment and the print-based library. The terminology used to describe converged services, which in some cases were co-located in one building, started to change to reflect their broader function. The 'learning centre' and its cousins were born, and job titles started to reflect the 'information' and 'learning' functions, rather than the place. Computing and library services were to be united in the provision of user-centred information services to learners.

In 1984 Battin, arguing from a USA perspective, stated that the way forward for libraries was to become 'scholarly information centers'[9]. This would be accomplished by merging the library and computer centre, and thus providing a 'one-stop-shopping' service for the university community capable of rapid response to fast changing technologies. Staffing of the 'scholarly information center' would be problematic in terms of funding, but 'trained subject specialists' and 'technical consultants' were envisaged, and such staff would need to be capable of providing a wide range of services to meet the information needs of a varied clientele, ranging from 'the freshman to the specialized scholar'.

Several years later, Cimbala referring to Battin's work, introduces the term 'hybrid' when exploring the need for the 'hybrid librarian-computer scientist' in the new electronic information environment[10]. Such a person would be capable of fulfilling both the technical and information skills roles in the modern electronic library. In 1992, Sutherland investigated the viability of the 'hybrid member of staff' within the UK, and came to the conclusion that although the idea had been popularised in early theoretical articles, it had neither materialised in practice nor been promoted as an ideal[11]. The consensus at the time was for the preservation of professional specialisms, with visits to institutions revealing a 'vertical career structure for most staff especially within the library'.

The eLib research project: SKIP – Skills for new Information Professionals, undertook an extensive interview programme with a variety of library and information staff in 1996 to 1997. SKIP found staff had mixed feelings about the viability and validity of hybrid staff[12]. There was concern that quality and standards were being compromised, and that traditional library skills were being 'watered down', as IT-based systems proliferated. Some participants expressed the view that librarians would suffer loss of identity as a result of diversification, whilst others felt it was simply impossible for one person to be expert in both information resources and information technology. Staff saw computing services and libraries as forming two distinct cultures, rather like the polarisation between the Arts and Science articulated by C.P. Snow[13]. Many participants did not see the need to gain new IT skills in order to survive, rather the opposite. They felt their professional identity was under threat from the mighty giant 'IT', against whom they had to arm themselves with the traditional tools and knowledge of librarianship. Some librarians saw IT as 'technical' and non-academic and somehow remote from the business of running libraries. It was okay to own a car and drive it, and perhaps check the dipstick once in a while, but what went on under the bonnet was strictly the domain of the car mechanic. A few staff had grown up with the DOS operating system or had a long association with computers. They were confident in the networked environment, and found that their skills gave them the advantage of being in great demand – although this did not necessarily manifest itself in terms of higher pay or respect from their peers.

In 1996 Sheila Corrall argued that professionals fell into one of two broadly defined groups: 'content professionals' – who specialise in information handling, and 'conduit professionals' – who are the IT or computing professionals[14]. Moves to conflate the two areas of expertise result in a dilution of 'expert contribution, at a time when our traditional areas of competence and strength ought to be pushing us into the limelight', she argues. This leaves one wondering why information professionals are being overlooked. Could it be that they are just too entrenched in the past and in 'library' procedures, and cannot see how to apply their expertise? Or is it a fault in perception which interprets any updating, supplementing, or pooling of skills as a dilution of expertise? An interesting development on this front involves the advertising of two posts at Reading University library, which is headed by Sheila Corrall:

> 'we are seeking two energetic, flexible and forward-looking professionals to fill new hybrid roles in our evolving innovative structure...'[15]

One of the posts requires the incumbent to produce printed and electronic guides. This requirement is in line with the definition of the hybrid library, which is a service delivering both print and electronic services. The term 'hybrid' here seems to mean someone with IT-based information skills, rather than computing or technical skills; someone who is able to work, equally well, with either electronic or traditional information resources. This advertisement does reveal that a paradigm shift has occurred since Sutherland undertook her research in the early 1990s. The hybrid is now being actively sought, and one assumes such staff are 'out there'– if posts such as these are being filled by staff who possess hybrid skills.

## The call for generalists

The call for hybrid staff has been matched more recently by a stated need for generalists. Phase 3 of the eLib Programme features five hybrid library projects which are now into their second year. The BUILDER project at the University of Birmingham is one of the five, and it has recently published a report based on data derived from a series of focus groups[16]. One of the topics under discussion was the roles and skills requirements of staff working in the hybrid library, and a consensus emerged that staff needed to be generalists rather than specialists. However, participants were realistic in voicing concern that staff did not have the diverse range of skills necessary for new roles in the hybrid library; roles which they identified as intermediary, trainer, and in quality assurance. However, some of the comments question the whole library approach to the new environment: 'We haven't rethought things. The electronic age has come and we are still doing the same things. We need to start from scratch.'[17]

The report notes that staff felt their futures were not assured. The traditional role of librarians was seen to be under threat, whilst an increasing need for information systems experts and administrators was identified: 'We need to develop a new role for ourselves or we will be by-passed'[18].

Despite fears of being left behind in the scramble for cyberspace, participants express total confidence in the printed word, and lack of faith in web-based information. They seem to feel that web-based resources should come with a health warning to users. This attitude to the printed word seems to reflect the longevity and familiarity of the book rather than the strict veracity of its contents. Admittedly, books purchased by academic libraries tend to be the revered works of scholars and experts in the field, but they offer an interpretation of events, or may even present a totally biased view. If higher education aims to teach students to question, analyse, and assess a variety of viewpoints, and to weigh up arguments and arrive at a conclusion of their own, then surely this is equally applicable to information found on the Internet? Yet the veracity of the printed word is not questioned, whilst electronic resources are feared and reviled by many library staff. Maybe this is the 'shock of the new' syndrome, due to the anarchic and commercial nature of the Internet. The web offers the end-user a wealth of instantaneous information in a format which is in tune with the visual, interactive, youth-oriented culture of today. Web-based information does not, as yet, come complete with title, publisher, author's name etc. and is therefore totally uncontrolled, and uncontrollable. It is perhaps this quality that librarians hate; the disorder, the vastness and the dynamic nature of the information. It cannot be pinned down like a butterfly specimen in a display case. However, web-based information is likely to play an increasing part in people's lives, as is the web as a medium for course delivery.

Once the thorny issue of metadata is resolved, and the world's electronic resources are as easy to locate as a book on a library catalogue, then librarians may feel differently about the Internet. This is unlikely to happen overnight, or to be the saviour of the profession. All library staff, irrespective of whether they operate in the public, private or academic domain, or whether they are 'professionals' or part of the growing band of non-library qualified staff, have to embrace the new technology, and see it not just as a tool but as an environment which will increasingly shape our lives. Education is on the political agenda, as the spate of recent government initiatives testifies. Lifelong learning; the National Grid for Learning and the University for Industry, are programmes which are designed to bring education and training to anyone who is able to benefit. ICT is seen as an enabler of learning, and organisations which successfully utilise its power, and bring appropriate education and training to the wider community, are likely to survive.

## The hybrid library

The hybrid library is a concept rather than an actual physical entity, although a physical library may be a constituent part of the hybrid library. The hybrid library is a cross-breed which aims to meet the needs of the new learning environment. It occupies both virtual space and physical space, and offers users access to a range of print and electronic resources. The following are definitions which originate from the eLib hybrid library projects:

> 'the hybrid library is on the continuum between the conventional and digital library, where electronic and paper-based information sources are used alongside each other' [19];
>
> 'the hybrid library is more than an interface offering integrated discovery of resources, whether they are in analog or digital format, whether they are local, national or international... the hybrid library is seen as an environment with physical and virtual services supporting professional activities of the users at their workplace from the discovery of information to the manipulation and analysis of the delivered resources'[20].

Fine words indeed and an excellent ideal, but can it be realised? The 'hybrid' library has its feet firmly planted in the soil of current practice – the traditional and tangible – but also has an eye fixed on the horizon of the 'foreseeable' future, where electronic information looms large. True to Darwinian principles, the hybrid library aims to survive in a fast changing world. It aims to meet the needs of the changing learning environment, which in turn aims to meet the needs of a changing society. The hybrid library provides information in any format – from traditional print to multimedia; access and holdings. It thus minimises the risk of redundancy. Neither the bath water nor the baby are thrown out – both are needed in times of drought or epidemic.

## The eLib hybrid library programme

The hybrid library, as a model for the delivery of current and future library and information services, is one of several areas within Phase 3 of the eLib Programme. Phase 3 aims to consolidate the learning of earlier projects by adopting four main approaches: Hybrid Libraries, Large Scale Resource Discovery, Digital Preservation and the development of early eLib projects into services[21].

The five eLib hybrid library projects each have a different approach or focus, for example: systems led; institutional and a range of client types. Chris Rusbridge, the eLib Programme Director states:

> 'the hybrid library was designed to bring a range of technologies from different sources together in the context of a working library, and also to begin to explore integrated systems and services in both the electronic and print environments'[22].

However, he sees the use of the name 'hybrid' as a reflection of the 'transitional state' of the library, and adds: 'which today can neither be fully print nor fully digital'. Compare this with the definition and explanation offered by the Agora Project (which takes its title from the Greek word for marketplace):

> 'The name, Agora, reflects the project's belief that the library will continue to organise the assembly places where information users and information products are brought into fruitful contact. Agora interprets the hybrid library as a wish to move to a service driven by user needs rather than by the constraints of location, mechanics of interaction, or media. The project believes that the importance of the open systems approach is that it moves the focus of attention from the technical challenges to issues of collection development, user needs, and the policy and practice of new libraries'.[23]

The HyLiFe Project – the Hybrid Library of the Future – involves six institutions of higher education. The present author is a member of the project team based at the University of Plymouth. The project is co-directed by the Information Services Department at the University of Northumbria at Newcastle, and the Centre for Research in Library and Information Management (CERLIM) at Manchester Metropolitan. The other partners are the University of the Highlands and Islands Project, the University of Central Lancashire, and the University of Newcastle upon Tyne.

HyLiFe is developing a series of user interfaces for different groups of users to provide access to hybrid library services. In the case of the Universities of Plymouth and the Highlands and Islands the users are based at a distance, and are geographically distributed. The implementation of the interface therefore has to look at how services are actually delivered and how they are supported. HyLiFe's primary focus is, therefore, with the human and organisational issues which underpin hybrid library services. The overall aim of the consortium is to establish, test, evaluate, and disseminate a knowledge of operating practices for the hybrid library within the higher education community.

Two of the main issues facing the HyLiFe Project at Plymouth is that of training and support for users of the interface, and the delivery of items users have identified when accessing the interface. Our users are nursing students who are undertaking courses at both pre- and post-registration level at two colleges which are at a distance from the main University campus. Both staff and students require hands-on training, and continuous support, before they are able to make good use of both electronic and print resources. They need easy and convenient access to computers if they are to establish a pattern of regular use. This problem is one which affects site-based library staff as well as students. If staff have limited access to computers they are unlikely to have the knowledge or skills to support students in the networked environment. Students may have access to dedicated computer suites, but they lack either the confidence or the necessary information and computing skills to make good use of these facilities. When time is precious, as it is for many of the nursing students who attend college on an irregular basis in between working on hospital wards, they resort to the local, the familiar and the portable. As Maurice Line states: 'the main thing is no-one wants to go to too much trouble'[24] – an observation which ought to be run up the library flag pole at regular intervals and saluted, especially where access to electronic databases is concerned, and users can end up needing a plethora of usernames and passwords to gain entry.

## Conclusion

The need to adapt, change, acquire new skills and roles, and update programmes of library education is convincingly argued in the literature – a small sample of which has been reviewed here. The warnings are loud and clear – 'adapt or die' – and there is now some evidence to suggest that they are being heeded, as job advertisements appear for staff with electronic information skills, and new courses are starting to be offered by departments of information studies[25]. The new breed of information professional required in the evolving learning environment is likely to have a mix of skills, but is more likely to be a generalist, rather than a true hybrid, who ideally would have advanced IT skills as well as information handling skills. Evolution is, however, a slow process, especially for long established, reactionary, hierarchical organisations and professions. Resistance to change cannot be underestimated, and in the electronic arena there are always others waiting in the wings, keen to replace those who fall by the wayside. As Blaise Cronin warns: 'the best of breed among the new players will develop sustainable and scaleable delivery strategies'[26]. The development of specialised user interfaces is a starting point, but to make the hybrid library a reality for the future you need to be able to back this up with reliable services. Services which can be 'scaled up' from the specific in-house project, which benefits a handful of users, to the institution as a whole, and is transferable to other institutions. Sustainability hinges on the availablity of funds and appropriate staff beyond the life of a project.

The hybrid library looks like a solution to some of the issues facing higher education. If universities take Cronin's advice they will embrace what he terms 'accelerated hybridicity' in order to keep pace with the 'transformational' pace of change[26]. Libraries would have to follow such a move, and this would entail no longer differentiating between campus based and distance learning. The seamless service, envisaged by the hybrid library projects, might therefore be extended to mean that users are not only provided with a uniform level of service, irrespective of information format, but are also treated equally irrespective of their mode of study.

To do this libraries will have to develop support systems in the form of online or telephone-based help services. Banks and supermarkets, which have diversified into new service areas, and are now in competition with each other, may provide some insight into the future for hybrid libraries. The call centre and Internet transactions, open up services to users when and where they want them. In the Learning Age the library may have to become more businesslike in its dealings with users; it may have to think in terms of diversification and marketing strategies. Now that education is no longer restricted to an elite, but is a mass programme and a lifelong process, academic libraries will have to adapt to survive. A new kind of thinking and mindset is needed, and this may involve a certain amount of risk taking. The hybrid library may also find it difficult to balance the needs of the old and the new, the print and the electronic, the traditional and the innovative. In the final analysis we are dealing with people, and in many cases it is lack of familiarity with, and a perceived threat from, IT systems which represent the biggest barrier to change. Training can only offer a partial solution; the individual, the organisation, and the information profession as a whole, has to be prepared to change and to adapt if it is to survive.

## References

1. *The New Encyclopaedia Britannica. Micropaedia Ready Reference*, Volume 6. (Chicago: University of Chicago, 1990, p.183.)

2. Beer, G. (ed.) *The origin of species by Charles Darwin.* (Oxford: Oxford University Press, 1996, ch. 4.)

3. A set of World Wide Web pages describing the Electronic Libraries Programme can be found at URL: *http://www.ukoln.ac.uk/services/elib*

*4. Joint Funding Councils' Libraries Review Group: report.* (Bristol: HEFCE, 1993.) (the Follett Report)

5. John Fielden Consultancy. *Supporting expansion: a study of human resource management in academic libraries.* (A report for the management sub-group of the *Library Review*). (Bristol: HEFCE, 1994.) (the Fielden Report)

6. Van Brakel, P. 'Education and training for information professionals in face of the Internet and World Wide Web.' In Raitt, D. (ed.) *Libraries for the new millennium: implications for managers*. (London: Library Association Publishing, 1997, pp.240-282.)

7. Van House, N. and Sutton, S.A. 'The panda syndrome: an ecology of LIS education.' *Journal of Education for Library and Information Science*, 37(2), 1996, pp.131-47.

8. Convergence is discussed in : *Relay: the Journal of the University College and Research Group* 42. See also: Collier, M. 'The context of convergence.' In Oldroyd, M. (ed.) *Staff development in academic libraries.* (London: Library Association Publishing, 1996, pp.68-79.)

9. Battin, P. 'The electronic library – a vision for the future.' *EDUCOM Bulletin*, 19(2), Summer 1984, pp.12-17, 34.

10. Cimbala, D. 'The scholarly information center: an organisational model.' *College and Research Libraries*, 48(5), September 1987, pp.393-398.

11. Sutherland, P. *The management of integrated learning resources.* (Brighton: Council of Polytechnic Librarians, 1992, pp.40-66.)

12. Garrod, P. and Sidgreaves, I. *Skills for new information professionals: the SKIP Project.* (London: LITC, 1998.) Also available at: *http://www.ukoln.ac.uk/services/elib/papers/other/skip*

13. Snow, C.P. *The two cultures; and, a second look*. new ed. (Cambridge: Cambridge University Press, 1993.)

14. Corrall, S. 'Defining professional competence: skills and prospects for the information profession.' Paper presented at:*Vision 2020: training for the new millenium. Circle of State Librarians Annual Study Conference,* London, 4 November 1996.

15. Email from Sheila Corrall to *lis-elib@mailbase.ac.uk* on 12 January 1999.

16. Hampson, A. BUILDER *The impact of the hybrid library on information services staff.* January 1999. *http://builder.bham.ac.uk*

17. Hampson, A. BUILDER *The impact of the hybrid library on information services staff.* January 1999. *http://builder.bham.ac.uk*, p.25.

18. Hampson, A. BUILDER *The impact of the hybrid library on information services staff.* January 1999. *http://builder.bham.ac.uk*, p.26.

19. Pinfield, S. *et al.* 'Realizing the hybrid library.' *D-Lib Magazine,* October 1998. *http://mirrored.ukoln.ac.uk/lis-journals/dlib/dlib/dlib/october98/10pinfield.html* p.1.

20. Pinfield, S. *et al.* 'Realizing the hybrid library.' *D-Lib Magazine,* October 1998. *http://mirrored.ukoln.ac.uk/lis-journals/dlib/dlib/dlib/october98/10pinfield.html*, p.3. (the MALIBU Project)

21. eLib Programme Phase 3 fact sheet, edition 4, p.3. Available at: *http://www.ukoln.ac.uk/services/elib/*

22. Rusbridge, C. 'Towards the hybrid library.' *D-Lib Magazine*, July/August 1998. *http://mirrored.ukoln.ac.uk/lis-journals/dlib/dlib/dlib/july98/rusbridge/07rusbridg*

23. Flyer for the Agora Project. See also: *http://hosted.ukoln.ac.uk/agora/*

24. Line, M. 'Designing libraries round human beings.' *Aslib Proceedings* 50(8), September 1998, pp.221-229.

25. Postgraduate course in: Electronic Information Management at Robert Gordon's University from September 1997.

26. Cronin, B. 'The electronic academy revisited.' *Aslib Proceedings* 50(9), October 1998, pp.241-254.

# Non-use of business libraries and information services: a study of the library and information managers' perception, experience and reaction to non-use

*Laura Brick*
*Assistant Librarian, Credit Suisse Financial Products, One Cabot Square, E14 4QJ*
*Email: laura.brick@csfp.co.uk*

*Not since 1980 has research provided any insights into how information managers perceive and cope with non-use. In an attempt to fill the vacuum information professionals in business companies were surveyed: 42 information managers were questioned and 12 interviewed. The study found that non-use is widespread, and that the non-use situation is a complicated mix of non-use and unsatisfactory user behaviour. The most commonly perceived cause of non-use was a lack of awareness of the services' capabilities. There is no stereotypical non-user. A significant majority of the managers had a positive attitude to non-use and were worried about the extent of non-users. Some tackled non-use head-on and implemented remedial treatment but less than a third could identify their non-users. Recent advances in information system design do not seem to have improved the situation and there has been little change in the scale of non-use, or reactions of information managers to it since 1980. The study concludes that the fruits of such an obviously positive attitude to tackling the non-user will only be gathered if it is converted into a reality and becomes an integral part of the information managers' marketing strategy.*

## Introduction

Many user studies have been conducted in the library and information field. This so-called user research has, on the whole, ignored the non-user. Indeed, the very term user would seem to exclude consideration of 'potential customers'. This paper aims to redress the balance somewhat by investigating non-use of commercial information services in the workplace by surveying and interviewing information managers about how they perceive and react to non-use.

This imbalance in user to non-user research seems unforgivable when one considers that for many years now the profession has been urged to adopt a modern marketing approach to their services[1]. It has been said that 'libraries are yet to play marketing games in the true sense'[2]. Yet perhaps they should, for as has been quite rightly stated, 'for users, library-information services may be the market leaders, for non-users they are obviously not'[3]. The Internet has made the market receptive to new ideas, making marketing services easier. And there are a lot of important questions that need addressing.

- Are librarians in the commercial sector taking advantage of a receptive market and now addressing non-use in order to extend their market?
- Have they been silently tackling the issue?
- What is the reality, nature and extent of non-use of internal business information services?

In the context of this study the terms 'library' and 'information service' are used interchangeably. The 'library/information manager' is the person designated to manage the resources of the unit. A 'non-user' is a person who might need information in the course of their work, has a staffed library available to them but do not use it, or use it in any of the contexts described below. 'Non-use' is defined as not using the information service provided in the workplace as an aid to work. It also, importantly, encompasses under-use or sub-optimal use, infrequent, superficial or inefficient use.

### Literature review/research context

The importance of studying non-use of libraries and the need to adjust user studies to include non-users was first voiced in the 1970s[4]. Indeed, in response to such calls for action there was increased research activity in the public[5] and academic[6] library fields. However, the findings of this research cannot be easily applied to commercial business information services. It was not until 1980 that any research was carried out specifically on non-use in the commercial workplace[7]. Very little of value has been conducted since. Slater's study of non-use of industrial and commercial information services represents the only British, or indeed international, attempt to examine non-use from the viewpoint of unit managers. Slater found that: 87% of her sample were aware that they had some kind of sub-optimal usage problem; 93% had encountered non-usage at some stage – past or present; and 70% were concerned to some extent about the existence of non-users. Positive specific remedies were suggested by 68% of respondents. A multiple approach to the solution of the non-use problem appeared to have the greatest chance of success. The most popular methods suggested were: organised self-promotion, less formal contact, education reform programmes and the demonstration of resources. Slater also produced the only published case study research of non-users of on-site industrial and commercial information units/libraries in Britain. She found that promotional activity, user induction and education was greatly needed within firms[8]. It was the intention of the author to replicate and update this very study.

Since Slater, there has been little research conducted in regard to non-use of libraries in the workplace. However, the issue has not been totally neglected in the professional press[9]. In 1994, a manager of an 'in-house' library wrote of her views, perceptions and experiences of non-users[10]. She suggested that, by studying non-use, managers might find ways to improve their service to all users[11]. In the same year Green explained that contributory factors to non-use included the images and perceptions people hold of what libraries are and what services they can offer[12] and that lack of awareness of services can form a barrier to use[13]. Other studies have found that non-users include: staff who are too busy to seek the help of the information staff; those who admit they need information but feel that they can manage without it; and those who believe their subject knowledge is sufficient to mean that they did not need library services[14].

### Methodology

The study was limited by time and resources to a study of business libraries only. Questionnaires, in combination with interviews, were considered the most appropriate method in order to glean quantifiable data of the scale of non-use and provide qualitative answers about manager's perceptions. The symbiotic nature of quantitative and qualitative methods was also taken into consideration. The sample of 80 for the questionnaires was selected at random from the membership list of a special interest group of one of the professional bodies. This at once brings into question how representative the sample was. Those bothering to join organisations may have more progressive attitudes and differ from the wider population of business information professionals from which they self-select themselves. This has to be admitted. In addition the sample was restricted to London based firms. Forty-two (53%) questionnaires were returned. The interviews were self-selecting from those that answered the questionnaire, and a total of 12 interviews were carried out.

### Results and discussion

The average size of the information units responding was between four and five people. The units with 11 or more staff were all well aware of non-use, so this might suggest that having more staff available does not reduce levels of non-use. Logic might also suggest that a larger budget would result in lower levels of non-use, because the extra money would enable the greater promotion of the service and better services to be offered. With only 45% of respondents answering the question on budgets no strong conclusions can be drawn; however, all the respondents with the biggest budget (£500,000 or above) experienced some level of non-use.

***User surveys***

Fifty-five per cent (23) of respondents did not carry out user surveys. The reasons for this varied from complacency: 'years ago I decided how I want the service and that people are happy'; to having alternative methods of measuring customer satisfaction, such as using 'the personal touch'. Of those that carried out user surveys, 11 targeted either everyone in the company or at least samples, which included non-users. The remaining eight distributed their user surveys to current customers only. So 26% of all the respondents did not take the term 'user' literally and surveyed all their potential customers.

***The scale of non-use***

It would seem that non-use is widespread. Eighty-eight per cent (37) of respondents were aware that they had some kind of sub-optimal usage problem (i.e. infrequent, superficial, last resort or inefficient usage). Only 12% (5) were unaware of any current problems in this area and only three of the 42 (7%) had never consciously encountered a non-usage 'problem'. If a comparison of these results is made with Slater's 1980 study the similarities are astounding. Eighty-seven percent of her sample was aware they had some kind of sub-optimal usage problem and, again, only 7% had never encountered a non-usage 'problem'[15]. It would appear that the situation was much as it was in 1980 – almost 20 years ago.

Respondents were also asked what percentage of staff in the organisation were users of the information service. The mode response was 31% – 40% and the mean was 50%, after the exclusion of the Management and Consultancy sector, which, as discussed below, had much higher levels of usage. These responses show that for many information managers levels of penetration was relatively low and, consequently, levels of non-use were high. By their own admission, on average 50% of what commercial information managers defined as the potential user population were in fact non-users.

***Sector differences***

There were some interesting results when the perceived non-use was viewed by financial sector. The mean percentage of staff who might be expected to use information at work and who were users of the service for the Management and Business Consultancy sector was 82%. For the Banking and Finance sector it was 46.8%. A third of the Management and Business Consultancy respondents indicated that there was no non-use, whereas no Banking and Finance respondents felt they were in this 'happy' position. The possible reason why incidences of non-use may be lower in the Management and Business Consultancy sector is that potential users are working in an information intensive field – libraries and their services cannot be ignored. As one information professional put it: 'the business we are in is very data driven and consultants would not be able to do their job without us'. Higher levels of non-use in the Banking and Finance sector can, perhaps, be attributed to cultural factors. One of the interviewees explained: 'The financial world is very much a closed society. We compete with other people and if you find out something you keep it to yourself.' (Banking and Finance combined service). Users are not accustomed to asking for information.

***The composition of non-use***

It was felt that the key to understanding the non-use issue was to examine in more detail the levels of non-use information managers were experiencing. Respondents were asked to indicate the nature of less-than-satisfactory usage patterns that they observed in connection with their services. Multiple answers were frequently given. The results are shown below in Table 1. When these results are compared to Slater's 1980 survey similarities can be noticed (Table 2)[16]. The omission of any reports of misuse and abuse of the service in the more recent survey can only be seen as a positive trend.

Analysis of these results suggests that 'non-use' is really a complicated mix of non-use and unsatisfactory or unrewarding user behaviour. The extreme form of the phenomenon – complete non-use – was experienced by a third of all respondents, which, although significant, was not seen as the most typical manifestation of the phenomenon. Minimal usage was a much more commonly encountered form.

### *Recent advances in information systems and the effect on non-use*

The interviewees felt that recent advances in information systems have led to changes in the nature of non-use. Use of the information centre directly was seen to have declined, with users becoming more self-sufficient. For example, one financial user (on receiving Reuters Business Briefing) stated: 'right that's it don't do any more searches for me'. There has also been a decline in the use of printed sources. These are, of course, not very surprising discoveries. However, interviewees felt that the situation was not as black and white as it appeared. Direct use of the information centre may be declining but employees were now using the desktop access to information systems provided for them by the information service. So in this sense it could be argued that use was rising.

*Table 1. Patterns of use and non-use.*

| **42** | **100%** | **Number of respondents** |
|---|---|---|
| 14 | 33% | Complete non-use observable amongst potential users |
| 24 | 57% | Minimal usage – infrequent and superficial |
| 18 | 43% | Frequent but sub-optimal usage |
| 5 | 12% | No apparent 'non-user' problems at all |

*Table 2. Patterns of use and non-use in Slater's 1980 survey.*

| **173** | **100%** | **Number of respondents** |
|---|---|---|
| 56 | 32% | Complete non-use observable amongst potential users |
| 112 | 65% | Minimal usage – infrequent and superficial |
| 71 | 41% | Frequent but sub-optimal usage |
| 22 | 13% | No apparent 'non-user' problems at all |
| 14 | 8% | Misuse and abuse (over-use, selfish delinquency) |

Has the Internet and other technological advances made people more information conscious and, as a result, more likely to use an information service? Possibly, but a majority (57%) of respondents with non-usage problems indicated that lack of information consciousness was still a cause of non-use. There was however overwhelming agreement with the attitudinal statement in the questionnaire: 'Now is a good time to tackle the problem of non-use because the general level of information consciousness in society is raised due to the Internet.' Also, all the interviewees felt that people have become more information conscious. At first glance this appears somewhat contradictory: information consciousness is perceived to be raised, but lack of information consciousness is given – by half of the respondents – as a reason for non-use. The answer may lie in the fact that managers felt that people in society and in general are more information conscious, but amongst the individuals who work in their company non-use is still a very real issue.

*Table 3. Characteristics of non-users.*

| **n** | **%** | **Characteristic** |
|---|---|---|
| 7 | 37 | Senior staff/director level |
| 4 | 21 | Older staff |
| 4 | 21 | Lazy/apathetic types |
| 3 | 16 | Research staff |
| 2 | 11 | New joiners |
| 2 | 11 | Support staff – secretaries |
| 2 | 11 | Consultants |
| 2 | 11 | Arrogant/complacent types |
| 1 | 5 | Surveyors |
| 19 | 100 | Total |

Respondents in Slater's 1980 survey had urged that available funds be put into a drive to raise the general level of information consciousness of the community, believing this would raise the status of the profession and the service it provides, and thus be the solution to the non-use problem[17]. The advent of the Internet has gone some way to achieving this but the results of the survey would suggest that it has not proved to be the solution to the non-use issue originally hoped for.

### ***Recognition of incipient or potential problems***

#### *Characteristics of 'non-users'*

Just over half the sample (51%) that experienced a 'non-usage' situation believed that certain categories of people were in fact predisposed to non-use. The cause may lie in their work context, or may be simply explained by personal characteristics (Table 3).

Senior staff would appear to be the main group of non-users, but the sample is too low to be certain about this. It could be that library managers are more aware of non-use by high profile staff. One respondent found that 12% of the total enquiries came from partners and another interviewee stated: 'Even though it isn't their job, some partners love information to pieces and come into the information centre.' (Banking and Finance information service). So certainly not all those at a senior level are guilty of non-use.

With nearly half the respondents who encountered a non-user situation believing that there were no objective group characteristics of non-users, and with the other half being so divided (with no one characteristic really dominant) no clear picture can be drawn of the non-user. Targeting the non-user thus becomes a complex and challenging proposition.

#### *Causes of non-use*

Respondents were asked why they thought 'non-users' made little or no-use of their services. Multiple answers were frequently furnished. Indeed, individuals often have multiple reasons for use or non-use of a library[18]. Results are shown in Table 4.

Lack of awareness of the service's capabilities was given as a possible cause by the majority (78%) of the sample. But, who is to blame for this situation? The managers interviewed were asked why they thought this was the case. One stated:

> 'The fact is it's all word of mouth. There's never been any formal promotion. We get queries from branches but usually they are just simple things. I don't think that they are aware that we've got a lot of back up to answer more complex queries.'
> (Banking and Finance combined service)

So some must share the blame and in these cases more promotion is obviously needed. However, others have promoted the service and its capabilities to no avail.

> 'We edit the company newsletter and we often have a page in there about what services we can offer. It's amazing though, people come along looking for a piece of information and we say 'yes we can provide you with that

*Table 4. Reasons for non-use.*

| **37** | **100%** | **Respondents after exclusion of those without non-users** |
|---|---|---|
| 29 | 78% | Lack of awareness of the service's capabilities |
| 24 | 65% | They delegate the search |
| 21 | 57% | They are not 'information conscious' |
| 15 | 41% | They prefer to use personal contacts |
| 14 | 38% | They believe their knowledge is sufficient |
| 12 | 32% | Location of the service, not close to potential users |
| 9 | 24% | Apathy on non-users' behalf |
| 7 | 19% | They are suffering from information overload |
| 2 | 5% | Low status of the service |
| 2 | 5% | Management attitudes to the service are non-supportive |
| 3 | 8% | Others |

didn't you read about it in the .....' They reply: 'No what was that? I don't remember seeing that."
(Banking and Finance combined service)

In the face of these experiences it was felt that you had to just keep on raising awareness and marketing the service.

That users delegate the search was the second most popular cause of non-use of the information service given by 65% (24) of the sample. This was not mentioned in Slater's earlier survey results but had been documented by Nicholas *et al.*[19]. In this survey it was felt users delegated the search to secretaries and office juniors; and partners delegated to managers and executives. This caused problems because 'it is harder to service their needs if they delegate the search'. (Insurance information service) Those interviewed felt it was important to keep in touch with the 'non-users' who delegate.

Nicholas *et al.'s* research, and Butcher's[20] also discovered that many in business believed in the benefits of personal contacts. This survey would appear to support these findings with 15 (41%) of the sample citing this as a reason for non-use. The use of contacts internally, which was described as 'effective networking not the 'old boy network'' (Banking and Finance combined service) by one manager, but 'use of the old boy network' by another (Accountancy combined service), is hard for the information service to compete with. However, the information managers interviewed did not see this as a threat and believed the Intranet and other 'knowledge management' systems could tap this knowledge.

On comparing Slater's survey results of 1980 with the author's a number of interesting findings emerge. Again lack of awareness of the service's capabilities was seen to be the most common cause of non-use (41% of Slater's sample named this as a cause). However, there was little support for two of Slater's causes – 'low status of the service' and 'management attitudes are not supportive', which were respectively supported by 29% and 26% of her sample. In the author's study only two (5%) respondents in each case indicated they perceived these to be causes of non-use[21]. Perhaps this suggests that an improvement in the status of business information services has occurred since 1980.

*Table 5. Did non-use worry the information manager?*

| 42 | 100% | Number of respondents |
|---|---|---|
| 19 | 45% | Yes, worried |
| 11 | 26% | Perhaps, slightly perturbed |
| 7 | 17% | No, not worried |
| 5 | 12% | No, no non-use problem so cannot be worried |

### *How do information managers react to non-use?*

All 12 interviewees agreed a level of non-use was probably inevitable. The majority of the respondents (71%) nevertheless were worried to some extent by the existence of non-users (Table 5).

*Table 6. Reaction of information managers who were worried about non-use.*

| n | % | Concern |
|---|---|---|
| 6 | 20 | They are not aware of what is available: 'we are not promoting' |
| 5 | 17 | Service not being used to optimum level: 'they undermine the service we provide' |
| 4 | 13 | Non-use lowering overall operating efficiency of the parent organisation |
| 4 | 13 | Survival fears: 'if they don't use us they may decide they don't need us' |
| 4 | 13 | They will set up end-user access to systems and by-pass the library/information centre |
| 3 | 10 | Non-users using inaccurate information |
| 3 | 10 | Non-users wasting their own time: 'reinventing the wheel' |
| 3 | 10 | Yes, worried but 'we could not handle any more enquiries' |
| 2 | 7 | Non-users are in senior positions who control our future |
| 30 | 100 | Total of respondents worried about non-use |

Even if we view these results as a percentage of the 80 in the original sample population it means at the very least 38% of information managers are concerned about non-use. A concern voiced by more than one manager in three should not be ignored.

Detailed reactions to non-use are presented in Tables 6 and 7. Multiple choices were made by some respondents. The most popular concerns do not appear to be sector specific and appear to be symptoms of professional insecurity and frustration that are probably common to all information managers. Certainly the most loudly voiced concern – lack of awareness – should be straightforward to deal with. A complacency brought on by high use or levels of work seems to be the main reason for a lack of concern about non-use.

*Table 7. Reaction of information managers who were not worried about non-use.*

| n | % | **Reasons for lack of concern** |
|---|---|---|
| 3 | 43 | Kept busy enough by the staff who do use the service |
| 2 | 29 | It is not my concern is people do not use the service |
| 1 | 14 | If all potential users became active users we would be swamped with enquiries |
| 1 | 14 | Library paid for by one department who account for 60% of enquiries |
| 7 | 100 | Total of respondents not worried about non-use |

### *A formal test of attitude and opinion*

A series of statements, representing various possible attitudes towards and opinions about non-use was presented to respondents in the questionnaire. People were asked where they stood in relation to the attitudinal stance of each statement, by ticking one of the following options: strongly agree, agree, disagree, strongly disagree and don't know. It enabled the collation of reactions to key concepts and issues from everybody who responded. The Likert scale, although having limitations[22], was chosen as the tool of analysis due to its ability to provide reliable, approximate ordering of people with regard to a particular attitude. In the case of this research a high score meant a positive attitude to the non-use issue. Therefore favourable statements such as 'efforts should be made to help any non-users who might benefit' were scored 1 for 'strongly disagree' and up to 5 for 'strongly agree'.

Individual total scores for each respondent were calculated. The possible range of scores was 70 (the highest possible score (5) x the number of statements (14)) to 14 (the lowest possible score (1) x the number of statements). The actual scores ranged from 40 to 62. All but three were above the midway point of 42 toward the positive end of the scale. The three not above the midway point scored 40, 41 and 42 so two were mildly negative towards the non-use issue and were on the 'neutral' midway point. The median score was 51.5 and the inter-quartile range was 49-56.

### *Reaction to statements concerning non-use*

The statements, which were used to produce the results on the Likert scale, are presented below. The order is that subjectively determined by respondents – ranging from strong agreement, through neutrality or mixed opinion to strong disagreement. To make for a comparison with Slater's survey results, answers were converted into scores. The system used was that: strongly agree answers were given a value of plus two; agree = plus one; don't know = zero; disagree = minus one and strongly disagree = minus two. The aggregate score for each statement is shown in Table 8.

The system worked well and did not conceal the reality of actual answer patterns, apart from where opinion was strongly divided. This happened only three times: 'one's primary duty is to the existing user group'; 'there are many more important and pressing problems than the non-user issue'; and 'non-use is a genuine problem, but in practice it's hard to do anything about it'. This was why the mode responses to each statement have also been included in Table 8.

The results displayed in Table 8 do seem to support the main finding of the Likert scale that managers of business information services were not indifferent to the non-user. They do feel very strongly that the non-user is a potential customer and that: 'efforts should be made to attract and help any non-user who might benefit'. This attitude was also very apparent 17 years earlier in Slater's survey. Where information managers disagree sharply with one another, both now and then, is about the chances of success in attracting the non-user[23].

*Do library managers identify non-users and discover why people are not using the service?* Of the 37 managers who confessed to having a non-user problem only 11 (30%) said they had made efforts to identify their non-users. The most popular method of doing this was through the personal approach – eight respondents said they did this. Non-users were targeted and made aware of the services offered. It was then felt that it was up to the individual concerned to choose whether or not to use the service, but it was felt that the information service had fulfilled its mission. Those who did approach and identify their non-users felt that their approach met with success. However, we must remember that over two-thirds of respondents did not identify their non-users.

*Remedial treatment*
Twenty-six (62%) of the information managers sampled believed that something could be done about non-use and suggested positive specific remedies. It was claimed that over 80% (21) of the time the remedial strategy, or some part of it, worked. There was no difference in methods used between those claiming successful non-user conversion and those making no such claims. It was stressed that

*Table 8. Levels of agreement with various attitudinal statements on non-use.*

| Score | Mode | Statement |
|---|---|---|
| +58 | Agree (24) | You need to know why people are not using your service |
| +55 | Strongly Agree (20) | Efforts should be made to attract and help any non-users who might benefit |
| +40 | Agree (30) | Studying non-use can allow you to market your service more effectively |
| +34 | Agree (26) | Now is a good time to tackle the problem of non-use because the general level of information consciousness in society is raised due to the Internet |
| +21 | Agree (24) | The cause of non-use is a state of mind (rather than external circumstances) so it should be possible and worthwhile to change non-user attitudes and behaviour |
| +13 | Agree (20) | Until better status within organisations is achieved by library/information services, usage will remain sub-optional |
| +4 | Agree 20 | There are many more important and pressing problems than the non-user issue |
| -4 | Agree 19 | One's primary duty is to the existing user group |
| -5 | Disagree 17 | Non-use is a genuine problem but in practice it is hard to do anything about it |
| -26 | Disagree 24 | Converting non-users would just stir up a demand that could not be met |
| -27 | Disagree (24) | Non-users are not sufficiently numerous to be worth worrying about |
| -33 | Disagree (23) | Changing the service to attract and accommodate non-users would result in a worse service to *current* users |
| -40 | Disagree (26) | The whole 'non-user problem' is just a theoretical fiction |
| -46 | Disagree (28) | Non-users are simply people who do not need information or documents to do their job |

any remedial treatment had to be ongoing. Also a multiple approach was favoured. The reasons for this were not made clear. Table 9 provides the details. Not surprisingly perhaps, given the foregoing, promotional activities and demonstrations figure most strongly. Promotion – used by 43% of respondents – took varied forms, from emails to posters. All, however, served the same purpose – that of detailing examples of enquiries answered and providing an explanation of resources available. As one interviewee stated: 'you have to market yourself like mad!'. (Insurance combined service.)

The use of presentations and demonstrations of the efficiency of the resources was used by 29% of the sample. One manager described how they set up live demonstrations with online databases. 'That's the best way – live demonstrations answering real needs. We ask them what they want to search and show them what we can do.' (Insurance information service.) This is interesting because, according to previous research, presentations do not feature highly in normal marketing strategies of managers of internal business libraries[24]. Perhaps, because they can be used to reach large numbers of potential users and can solve what was deemed to be the main cause of non-use – lack of awareness of the services capabilities – they were thought beneficial when tackling the non-user.

What is surprising about the survey results though is that friendliness and personal contact were not rated more highly.

In Slater's 1980 survey, promotion was again the most popular method said to be used to tackle non-use and was employed by 30% of her sample. Personal contacts and friendliness was the next most popular method, favoured by 29%, and followed by education and demonstrations[25]. It would appear that the most popular methods in 1980 remained the most popular methods in 1997. Novel or revolutionary solutions are absent.

## Conclusions

Given the size of the sample and its geographical limitations the findings of this study can be nothing more than indicative. However, it does appear that there has been little change in the level of non-use and reactions and attitudes of managers to it since 1980, when the only other published study of this topic was carried out. Non-use appears to be widespread amongst information services in the business sector. Thirty-seven of the 42 respondents (88%) were aware they had some kind of sub-optimal usage problem. Non-use is a complicated mix of non-use and unsatisfactory or unrewarding user behaviour. Although a third of respondents were aware of complete non-use in their organisation, the most commonly encountered form of non-use was minimal usage. This form of non-use may be more difficult to combat, for in this instance drawing the existence of the service to their attention is unlikely to bring forth dividends.

The size of department or budget seem to make little difference as to the size of the non-use problem, with heavily staffed and well-resourced departments still experiencing non-use. There were, however, sector differences between the two main sectors involved in the survey. The Management Consultancy sector showed high levels of usage, whereas the Banking and Finance sector showed above

*Table 9. Remedial strategies suggested by information managers.*

| n | % | Strategy suggested |
|---|---|---|
| 18 | 43 | Promotion publications, posters, leaflets etc. |
| 12 | 29 | Presentations/making the information tangible/demonstrations |
| 8 | 19 | Personal contact/friendliness |
| 7 | 17 | Identify and target non-users/ targeted services/SDIs |
| 6 | 14 | Target new recruits |
| 4 | 10 | Attendance at meetings |
| 3 | 7 | Word of mouth/referral market |
| 3 | 7 | Offering a quality service |
| 2 | 5 | Guided tours |
| 1 | 2 | System redesign end-user access |
| 42 | 100 | Total number of respondents |

average levels of non-use amongst potential users. A possible explanation for this is the difference in culture and any future study of this topic would do well to concentrate on the Banking and Finance sector to try and learn more about reasons for non-use. Just over half of those information managers who admitted to a non-use problem felt that there was no non-user stereotype.

Findings concerning the causes of non-use are partly supported by other researchers. The main reason suggested in this research – lack of awareness of the service's capabilities – had not however been shown to be a major cause in previous research, with only Green finding this to be the case[26]. If this really is the main cause of non-use of business information services in the workplace it must surely be interpreted as good news for the profession, for this is surely something that is easily tackled by a systematic marketing strategy.

It would seem that non-use is perceived as a problem, or at least undesirable. The majority of the information managers were worried to some extent by the existence of non-users. A positive attitude to the non-user situation was also suggested by these findings. The formal test of attitude found all but three respondents had a positive attitude towards the non-use issue. It would appear the majority of the profession is united in its concern about non-use.

Some insights have been gained into how library managers react to and cope with the non-user. Generally it was felt that now was a good time to tackle the issue because of the advances in information system design. In terms of remedial treatment there was strong support for self-promotion and the use of presentations to non-users. It was felt that there was not one miracle cure and a multiple on-going approach was preferred. However, this evidence does not mean the evolution of a new breed of proactive manager, for although a need to understand and identify the non-user has been advocated in the professional press[27,28,29], as yet managers appear not to be heeding the advice. For the majority (70%) no systematic effort has gone into understanding and identifying the non-user, even though this was given a lot of lip service.

This study has hopefully fulfilled its fundamental aim of producing up-to-date research in a neglected field of study. The message that should be taken away from this study is that a positive attitude to an issue will come to nothing unless it is transformed into decisive action. The fruits of such a positive attitude to non-use will only be gathered if it is converted into a reality, and the identification and targeting of the non-user becomes an integral part of the information managers' marketing strategy.

## References

1. Yates-Mercer, P. and Steward, Y. 'The Marketing of internal business information services.' *Journal of Information Science* 17, 1991, pp.221-233.

2. Sridhar, M. 'Non-use and non-users of libraries.' *Library Science* 31(3), 1994, p.115.

3. Sridhar, M. 'Non-use and non-users of libraries.' *Library Science* 31(3), 1994, p.115.

4. Klintoe, K. 'User research and its applications – current problems.' *Eurim II*, 1977, pp.157-158.

5. Whatley, A. 'The untapped market for the public library: a survey.' *Library Association Record* 80(9), 1978, pp.447-467.

6. Cannon, P. 'Non-users of academic libraries.' *Reference Librarian* 31, 1990, pp.119-129.

7. Slater, M. *The neglected resource: non-usage of library information services in industry and commerce.* (London: Aslib, 1981.)

8. Slater, M. *Non-use of library information resources at the workplace: a comparative survey of users and non-users of onsite industrial-commercial services.* (London: Aslib, 1984.)

9. Zhang, X. 'Reflections on non-users.' *Library and Information Service* 3, 1987, pp.18-19.

10. McCarthy, G. 'Getting to know your non-users.' *Library Management* 15(4), 1994, pp.30-34.

11. McCarthy, G. 'Getting to know your non-users.' *Library Management* 15(4), 1994, p.33.

12. Green, T. 'Images and perceptions as barriers to the use of library staff and services.' *New Library World* 95(1117), 1994, pp.19-24.

13. Green, T. 'Images and perceptions as barriers to the use of library staff and services.' *New Library World* 95(1117), 1994, p.21.

14. Buryi-Shmar'yan, O. 'The problem of information 'non-users': can it be solved?' *Nauchno-Teknicheskaya Informatsiya* 1(10), 1984, pp.13-21.

15. Slater, M. *The neglected resource: non-usage of library information services in industry and commerce.* (London: Aslib, 1981. p.27.)

16. Slater, M. *The neglected resource: non-usage of library information services in industry and commerce.* (London: Aslib, 1981, p.29.)

17. Slater, M. *The neglected resource: non-usage of library information services in industry and commerce.* (London: Aslib, 1981, p.11.)

18. Green, T. 'Images and perceptions as barriers to the use of library staff and services.' *New Library World* 95(1117), 1994, p.23.

19. Nicholas, D., Harris, K. and Erbach, G. *Online searching. Its impact on information users*. (London: Mansell, 1987, p.128.)

20. Butcher, H. 'Why don't managers use information?' *Managing Information* 4(40), May 1997, p.30.

21. Slater, M. *The neglected resource: non-usage of library information services in industry and commerce.* (London: Aslib, 1981. p.31.)

22. Oppenheim, A. *Questionnaire design, interviewing and attitude measurement*. new ed. (London: Pinter Publishers, 1992, p.200.)

23. Slater, M. *The neglected resource: non usage of library information services in industry and commerce*. (London: Aslib, 1981, p.35.)

24. Yates-Mercer, P. and Steward, Y. 'The marketing of internal business information services.' *Journal of Information Science* 17, 1991, pp.228.

25. Slater, M. *The neglected resource: non-usage of library information services in industry and commerce*. (London: Aslib, 1981, p.38.)

26. Green, T. 'Images and perceptions as barriers to the use of library staff and services.' *New Library World* 95(1117), 1994, p.20.

27. McCarthy, G. 'Getting to know your non-users.' *Library Management* 15(4), 1994, pp.30-34.

28. Sridhar, M. 'Non-use and non-users of libraries.' *Library Science* 31(3), 1994, p.128.

29. Butcher, H. 'Why don't managers use information?' *Managing Information* 4(40), May 1997, p.30.

# Constructing 'users' in library and information science

*Heidi Julien*
*School of Library & Information Studies, Faculty of Management, Dalhousie University, Halifax, Nova Scotia, Canada B3H 3J5*
*Email: hjulien@is.dal.ca*

**Abstract** *This article issues a challenge to information service providers and to researchers in information management to reflect on the ways the concept of 'users' has been constructed and applied. It is argued that we need to be mindful of the conceptual parameters we establish by our notions of users and the implications of those parameters as these inform our attitudes and endeavours. Alternate labels are considered but found wanting. Thus, the author puts out a call for a worthy replacement that can enlarge the scope of our understanding and help us to reconstruct our thinking.*

## Introduction

Ubiquitous in the literature of library and information studies is the term 'user', generally employed to refer to clients of information services and libraries, and to participants or respondents in research studies. It is time that we stepped back to examine our application of the term user and the implications arising therefrom, in as much as language constructs our understanding, influences our attitudes and behaviour, and establishes the parameters of our vision.

The Oxford English Dictionary (OED) defines 'user' as:

a) 'one who has or makes use of a thing; one who uses or employs anything';
b) 'a person who takes narcotic, etc., drugs';
c) 'a person or organisation that makes use of a computer'[1].

There is also another meaning not mentioned in the OED, that is, to exploit others for one's own advantage or gain. While it is obvious that information professionals generally do not employ user in either of the pejorative senses, that is, to refer to clients as drug users or as exploiters, these meanings remain unavoidably attached to the term. Clearly, it is in the contexts in which people make use of information services, and more specifically, of computers, that we normally employ the term user.

## Why only users?

Applying the term users naturally limits the scope of our interest. People who use formal information services, such as libraries, are a self-selected population that finds these services useful, or members of some group that is captive in some respect (e.g., students or scholars using a university library, or employees of an organisation). What of information needs not typically met by formal information services? Are these within the boundaries of our interest? Do we mean to exclude non-users from consideration? One would imagine that as long as complete market saturation is unachieved for a particular information service, then non-users would be of considerable interest. On a theoretical level, if we exclude non-users of formal information services from our focus, how are we able fully to explore human information behaviour? For it is only with understanding of the complexities of information behaviour that useful information services can be developed. Surely if such theoretical understanding may be translated into the design of improved information services, those of us with a vested interest in maintaining our professional turf ought to take note of the behaviour of those who currently choose means of satisfying their information needs outside of existing information services.

This point about the potential importance of the non-user is often poorly understood. One of the field's prominent proponents of attention to information behaviour, Diane Nahl[2], suggests an equivalency between user-centred and human-centred or people-centred approaches. Her lack of a distinction between 'users' and 'human's or 'people' effectively defines 'humans' and 'people' as users, which is a logical fallacy. This clearly is what occurs in Nahl's research, which concentrates solely on actual users of information systems and services. Ultimately, this focus perpetuates the vilified 'systems-centred' approach to our field, which has been widely touted as having been replaced by a 'user-centred' approach. A sober analysis of our user-centredness reveals that what we continue to be particularly interested in is the user of information systems, with scant attention to the non-user of formal information services. Thus, user-centredness is essentially an approach not unlike systems-centredness. The utility of the 'system' is the heart of our concern, not the information behaviour of people (including non-users), nor the interests of the user. Of course, there are a few notable exceptions to our systems-focus, i.e., research that analyses the actual information needs and behaviour of people, without necessary reference to existing, formal information services or systems[3].

### Users as inadequate half-wits

Even with the best of intentions, however, our application of the term user to refer to persons who use information services, or who use computers, is problematic. Our concept of users is remarkably unflattering. Examining our construction of users in the literature of the field reveals that we typically envision users as bungling fools who 'lack' appropriate skills to effectively find what they are seeking, and whose affective behaviours are at best only an annoying interference with effective application of cognitive skills to information retrieval, but which, at worst, are the primary barriers to effective information retrieval[4]. Thus, we limit our interests to the rational, cognitive, aspects of information behaviour, choosing to ignore or to denigrate the role of affect in information behaviour and in decision-making. Our self-imposed limitation blinds us to the myriad emotional variables that influence information behaviour.

In addition, we conceive of users as incompetents or as 'patients' whose symptoms require diagnosis. We construct ourselves (i.e., information managers or service providers) as experts with expert information solutions. If information seekers find our systems, digital or otherwise, impenetrable, we focus on instructing in adequate use, instead of re-designing systems to suit the information behaviours of information seekers.

In constructing our positions as experts and our clients' positions as novices who require help, we shape an unequal power relationship. In Western societies, accepting help has connotations for the recipient of 'inferiority, dependency, and inadequacy'[5]. Therefore, constructing clients as inadequate, whether their inadequacy arises from what we have termed as 'gaps', or 'anomalous states of knowledge', for example, places these hapless folk in a state of dependency. A classic example of this 'expert', and ultimately systems-centred, attitude is apparent in Allen's discussion of research showing the value of social networks in meeting a need for parenting information. He states:

> 'The clear implication is that the social network of friends and family acts either to diminish the perceived need for information or to decrease the incentive people have to obtain information about parenting.'[6]

What is not recognised is that whatever the needs of the research participants were, these were met by their social networks. Information sources that are outside our formal, expert information systems are not acknowledged as legitimate. Allen also describes information seekers as having perceptual 'failures', or 'inadequate' knowledge structures, or 'inappropriate or ambiguous social signals'[7]. He complains that people 'interact with their environment by limiting their intake of information, ignoring information if it is associated with negative outcomes, and taking information shortcuts'[8]. In all instances, the terminology selected acts to construct information seekers as deficient or inadequate. It is quite possible that from the perspectives of information seekers, limiting information intake, ignoring some information, and seeking efficiencies in information seeking are all adaptive behaviours, associated with positive outcomes. The parent who seeks help to cope effectively in this role may quite legitimately find that the 'information' provided by a supportive social network may efficiently meet affective, as well as more instrumental, cognitive needs. 'Expert' information provided through a printed or online source, more likely to be viewed by information managers as authoritative and therefore more trustworthy, is likely to be viewed by the parent as significantly more difficult to locate and interpret, as less trustworthy, and as providing little affective support.

### Users as elites

In addition, when we speak of users, we continue to focus on elite groups. Our research endeavours still tend to concentrate on the high-status user, as we attend to a restricted range of information needs, largely in terms of specific data and documents required by elites. A recent content analysis of research into information needs and uses demonstrated that the identifiable group studied most frequently in information needs and uses research is scholars[9]. And, the users of academic, public, and corporate libraries or information services, tend to represent society's elite social strata. Naturally, our preoccupations only reflect social power structures, as well as the interests on the part of research funding bodies. Nevertheless, do we really wish to confine the range of our concerns so narrowly? How often do we ask ourselves: in whose interests are we undertaking research?

### Where to from here?

So, where to from here? With apologies to Socrates, the unexamined term is not worth using. If user is unacceptable, what is left? Unlike those associated with user, there are no parallel deprecatory associations with the terms 'client', 'customer', or 'patron'. However, each of these terms is problematic in its' own right. Client suggests a particular type of professional relationship that is not universally endorsed in information management[10], and, like customer or patron, continues to exclude from consideration those individuals who do not interact with a formal information system or source. Customer is associated with a business model of service provision that remains controversial[11]. It is a term worthy of serious consideration, however. Evoking notions of financial transactions, employing 'customer' provides an opportunity to consider the concepts of cost and pricing beyond the purely financial. Remembering that those whom we serve pay psychological, social, and convenience costs can only spur improvement in service delivery. Patron is clearly outmoded, eliciting images of wealthy benefactors and guardians, and having no discernible place in the online world. Some time ago, Michael Gorman suggested a competition for a new term to describe information seekers who make use of libraries; he proposed the hideous term 'libraree'[12]. Sixteen years later, we are still seeking a term that meets the following conditions:

1. is inclusive; i.e., refers to information seekers who use formal information systems, as well as those who choose not to do so (which eliminates 'user', 'client', 'customer' and even narrower terms such as 'reader' or 'borrower'); and
2. applies to the widest possible range of formal information sources (which eliminates 'libraree').

Beyond attention to terminology, our conceptions of information seekers must be broadened to include non-elites, and to include those aspects of information behaviour that are observable in informal contexts. Theoretical approaches to information behaviour can no longer be limited to cognitive aspects. As Brenda Dervin has been telling us for years: it is nonsensical to conceive of people's minds as open buckets, into which we can toss 'information' bricks.[13] People use 'information', in whatever form, from whatever source, to construct, within existing cognitive and affective frameworks, some 'sense' or understanding. If we really are in the information game, then we must recognise that information comes in a wide variety of forms and from a wide variety of sources. In some contexts, particularly those that we as information professionals tend to inhabit, information is fully commodified, and treated very much like a set of bricks. In a broader sense, and certainly in other, equally legitimate contexts, information is anything that helps people make sense of themselves, and their worlds. If we really do want to play a meaningful role in people's construction of personal meanings, then our conceptions of what constitutes legitimate information, and of who is deserving of theoretical attention, must be broadened. Rather than imposing an identity on people seeking information to construct meaning in their lives, we need to allow, in a theoretical sense, and in our attempts to facilitate that meaning-making, for people's self-construction of meaning and behaviour[2,4,14].

## Conclusion and challenge

It is time for a thorough examination of the parameters we have established around those theoretical concerns we deem worthy of attention, of the practical, service implications of those parameters, and of the implications of our constructions for information seekers. People who use the systems we provide and promote deserve such an investment, and people who find our systems unfriendly, unhelpful, or simply immaterial, deserve it too. Surely those of us with particular concern for information behaviour and for information management collectively are able to recognise the inadequacies of our current terminology and conceive of a worthy alternative. Can we start a new millennium with a new term denoting a modified, more inclusive and more sympathetic, conception of our interests?

## References

1. *Oxford English Dictionary*, 2nd ed. (Oxford: Clarendon Press, 1989, pp.356-357.)

2. Cf. Nahl, D. 'An integrated theory of information behavior: taxonomic, psychodynamic, ethnomethodological.' *http://www2.hawaii.edu/slis.old2/nahl/articles/integrated/integrated/ http://www2.hawaii.edu/~nahl/articles/integrated/integrated_ToC.htm* (visited15 March 1999); Nahl, D. 'The user-centered revolution: 1970-1995.' *http://www2.hawaii.edu.slis.old2/nahl/articles/user/user1toend_toc.html http://www2.hawaii.edu/~nahl/articles/user/user1toend_toc.html* (visited 15 March 1999) Also published in *Encyclopedia of Microcomputers* 19 (New York: Marcel Dekker, Inc., 1996, pp.143-199.)

3. Cf. Chen, C. and Hernon, P. *Information seeking: assessing and anticipating user needs*. (New York: Neal-Schuman, 1982.) Chatman, E.A. *The information world of retired women*. (New York: Greenwood Press, 1992.)

4. Tuominen, K. 'User centered discourse: an analysis of the subject positions of the user and the librarian.' *Library Quarterly* 76(4), 1997, pp.350-371.

5. Nadler, A. 'Help-seeking behavior as a coping resource.' In Rosenbaum, M. (ed.) *Learned resourcefulness: on coping skills, self-control, and adaptive behavior*. (New York: Springer Publishing Company, 1990, p.129.)

6. Allen, Bryce L. *Information tasks: toward a user-centered approach to information systems*. (San Diego: Academic Press, 1996, p.39.)

7. Allen, Bryce L. *Information tasks: toward a user-centered approach to information systems*. (San Diego: Academic Press, 1996, p.109.)

8. Allen, Bryce L. *Information tasks: toward a user-centered approach to information systems*. (San Diego: Academic Press, 1996, p.109.)

9. Julien, H. 'A content analysis of the recent information needs and uses literature.' *Library and Information Science Research* 18, 1996, pp.53-65.

10. Gorman, M. 'A borrower is a client is a patron is a user is a reader.' *American Libraries* 14(9), October 1983, p.597.

11. Jones, B. 'Customers are consumers of library resources and services – or are they?' *The Australian Library Journal* 47(2), May 1998, pp.131-144.

12. Gorman, M. 'A borrower is a client is a patron is a user is a reader.' *American Libraries* 14(9), October 1983, p.597.

13. Dervin, B. *An overview of sense-making research: concepts, methods and results to date.* (Seattle: University of Washington School of Communications, 1983.)

14. Talja, S. 'Constituting 'information' and 'user' as research objects: a theory of knowledge formations as an alternative to the information man-theory.' In Vakkari, P., Savolainen, R. and Dervin, B. (eds.) *Information seeking in context.* (London: Taylor Graham, 1997, pp.81-96.)

**Subscriptions**
*Aslib Proceedings* is published ten times per annum and supplied to corporate members of Aslib. For 1999 the annual subscription rates are £170/$280 (£136/$224 to members of Aslib). For subscriptions and back issues please contact: Portland Press Ltd, Commerce Way, Whitehall Industrial Estate, Colchester CO2 8HP, United Kingdom. Tel: +44 (0) 1206 796 351. Fax: +44 (0) 1206 799 331.

---

Aslib, The Association for Information Management, is a world class corporate membership organisation with over 2000 members in some 70 countries. Aslib actively promotes best practice in the management of information resources. It lobbies on all aspects of the management of, and legislation concerning, information at local, national and international levels. Aslib provides consultancy and information services, professional development training, conferences, specialist recruitment, Internet products, and publishes primary and secondary journals, conference proceedings, directories and monographs.

Further information can be obtained from: Aslib, The Association for Information Management, Staple Hall, Stone House Court, London EC3A 7PB, UK. Tel: +44 (0) 171 903 0000. Fax: +44 (0) 171 903 0011. Email: aslib@aslib.co.uk *http://www.aslib. co.uk/*

---

Periodicals Postage Paid at Rahway NJ and at additional mailing offices.
**POSTMASTER:** Send address changes to PROCEEDINGS, c/o Mercury International, 365 Blair Road, Avenel NJ 07001

---

©1999 Aslib and contributors ISSN 0001 253X

ISSN 0001 253X
Volume 51
Number 8
September 1999

28/2/2011
wcl

Aslib

# Proceedings

## New information perspectives

Perspectives on information overload
David Bawden, Clive Holtham
and Nigel Courtney

Derivation of interaction environment and information object properties for enhanced integrated access and value-adding to electronic documents
Chern Li Liew and Schubert Foo

The introduction of intranets into the newspaper industry
Jackie Drennan

Building the digital library on solid research foundations
Ian Rowlands and David Bawden

Aslib THE ASSOCIATION FOR INFORMATION MANAGEMENT

*Aslib Proceedings* is a long-standing professional journal of international repute, which has a large and wide subscriber base. The journal brings currency, authority and accessibility to the reporting of current research, issues and debates in the broad area of information work. Above all the journal wishes to provide research and comment in a form that is easily and quickly understood: a fresh, rigorous, but unfussy, writing style is what is aimed for. The needs of busy practitioners (and academics) are very much in mind. Articles are refereed to meet the need for accredited and authoritative information. To ensure that this does not result in an unacceptable loss of currency, articles are speedily refereed by the journal's large, eminent and multi-disciplinary Editorial Board (assembled for just such a purpose). In exceptional cases they may be referred to an external panel of consultants. The journal's monthly publishing pattern helps in maintaining the currency of reporting. In consequence the submission of research in progress articles is encouraged.

The boundaries of information work are rapidly changing and widening. *Aslib Proceedings* has a role to play in the repositioning of the information profession by encouraging submissions from information managers, providers, packagers and consumers of all kinds – and especially those from the related information fields of journalism, electronic publishing, communication and Internet studies. An interdisciplinary Editorial Board has been assembled to assist in the process.

Prospective authors should submit articles of 2,000-6,000 words in length. Articles should be accompanied by a 100-150 word informative abstract and the numerical form of referencing should be adopted. In style and tone, submissions should fall between that of the journal's sister publications: *Managing Information* and *Journal of Documentation*. Articles should be furnished by email or on disc to David Nicholas, Editor, Aslib Proceedings, Department of Information Science, City University, Northampton Square, London EC1V 0HB. Email: nicky@is.city.ac.uk. The Editor is happy to provide advice on an informal basis to potential authors.

**Editorial Board**
Dr David Nicholas, *Head of Department, Department of Information Science, City University;* John Akeroyd, *Director, Learning Resources, South Bank University;* Sarah Blair, *Head of Publications, Aslib;* Peter Chapman, *Information Manager, NewQuest North East;* Professor Peter Cole, *Head of Department, School of Journalism, University of Central Lancashire;* Dr Tom Dobrowolski, *Senior Lecturer, Institute of Information Science and the Book, University of Warsaw;* Harry East, *Senior Research Fellow, Department of Information Science, City University;* Mike Hyman, *Managing Director, Head Software International;* Robin Hunt, *Creative Director, AreHause;* Nat Lievesley, *Network Manager, Centre for Policy on Ageing;* Helen Martin, *Information Manager, Research and Information Unit, The Guardian*; Dr Ian Rowlands, *Lecturer, Department of Information Science, City University;* Richard Withey, *Director, New Media, News International;* Ralph Adam, *Freelance Journalist.*

**Editorial Team**
*Editor:* David Nicholas
*Deputy Editor:* Ian Rowlands
*Editorial Manager:* Kay Nicholas
*Editorial Assistants:* Peter Williams (City University), Colette Coles (City University)

# CONTENTS

249-255 Perspectives on information overload
David Bawden, Clive Holtham and Nigel Courtney

256-268 Derivation of interaction environment and information object properties for enhanced integrated access and value-adding to electronic documents
Chern Li Liew and Schubert Foo

269-274 The introduction of intranets into the newspaper industry
Jackie Drennan

275-282 Building the digital library on solid research foundations
Ian Rowlands and David Bawden

# Perspectives on information overload

*David Bawden*

*Department of Information Science, City University, Northampton Square, London EC1V 0HB*
*Email: dbawden@is.city.ac.uk*

*Clive Holtham and Nigel Courtney*

*City University Business School, City University, Northampton Square, London EC1V 0HB*

**Abstract** *Information overload is by no means a new concept, but has come to prominence during the last decade. This paper reviews the nature and causes of overload, and considers possible solutions, both organisational and technical, and its relevance to the information professional.*

## Information overload; meaning and history

There is no single generally accepted definition of information overload. The term is usually taken to represent a state of affairs where an individual's efficiency in using information in their work is hampered by the amount of relevant, and potentially useful, information available to them. The information must be of some potential value, and it must be accessible. It is usually associated with a loss of control over the situation, and sometimes with feelings of being overwhelmed. In the extreme, it can lead to damage to health. We sum this up by suggesting that:

> information overload occurs when information received becomes a hindrance rather than a help when the information is potentially useful.

The concept of information overload is by no means new. The writer of *Ecclesiastes*, who remarked that 'of making many books there is no end; and much study is a weariness of the flesh' (ch 12, v 6), was the first of a long line of commentators who saw the proliferation of information as a detriment to effectiveness and efficiency.

Innovations in information technology, such as the printed book, the periodical magazine or journal, the abstracting journal, and the computer, have all led to complaints that it is impossible to keep up with the amount of information available. Such complaints have increased steadily over time.

By the late 1950s and early 1960s, with an exponential expansion of publication, particularly in science and technology, and the increasing take-up of mechanised documentation and computerised information handling, information overload was generally accepted as a problem:

> 'Torrents and rivers of current literature pour themselves into libraries, adding, without cease, to what is already there .... The scientist's time and power of attention are precious things which need to be husbanded; to do this we need techniques of controlled selectiveness in supplying his needs.'[1]

By 1973, Torkelson[2] was advocating the creation and operation of specific information services, in this case selective dissemination of information (SDI), as a means of overcoming information overload.

By the 1990s, information overload began to be referred to as a major problem, particularly with the influence of new technologies, particularly electronic mail and the Internet. This was crystallised by a series of reports of large-scale surveys[3,4,5], which indicated that information overload was a major problem for individuals and for organisations, and that the techniques and 'coping strategies' used in the past were no longer effective. The problem was affecting the effectiveness, and even the health, of professional workers, particularly managers in businesses, and was severely affecting the efficient working, and productivity, of organisations. The Reuters[3] survey of business managers, *Dying for information,* revealed a number of startling statistics:

- two thirds of managers believed information overload had caused loss of job satisfaction;
- two thirds believed it had damaged their personal relationships;
- one third believed it had damaged their health;
- nearly half believed important decisions were delayed and adversely affected as a result of having too much information.

In specific working environments, Wilson[6] has suggested that relevant information known to be available may go unused in research and development because of information overload, which may be 'a sign of strategic error', or may be accepted as 'routine and normal', while Hunt and Newman[7] note concerns about medical 'knowledge overload' in the use of various information sources by US medical practitioners.

An insight into current perceptions of overload among a relatively sophisticated group of information users is given by part of a survey of the use of information systems and services in a pharmaceutical research organisation[8]. One aspect of the survey related specifically to perceptions of information overload. When 30 scientists were asked to agree or disagree with the statement 'I have more information than I can cope with', 17 'agreed' or 'strongly agreed', but 8 'disagreed' and 5 were 'neutral'. Several of those who disagreed, or remained neutral, qualified their answer by saying that there was too much information available, but they had found ways to cope. Information overload therefore seemed to affect the great majority of the participants.

When prompted, most replied that IT in part caused the problem, but had, to some extent, provided the tools to help solve it; it was seen very much as a 'two edged sword'. The participants were evenly split as to whether the situation was better, worse or the same as five years ago, a small majority favouring 'worse'. Three participants refused to complain, on the grounds that the problem of five years ago was not enough information.

Two factors seemed to be involved. The first was the greater amount of information available, both internally and externally (associated with the greater complexity of most participants' jobs), so that ever more selection and discrimination was needed in dealing with it, when it is actively sought. The second was the amount of information sent directly to the participants, with email universally seen as the main source.

The solution was largely seen to be self-discipline in focusing on important things, and self-education in managing information. Several participants saw scope for formal training in managing their own information: file handling, what to store and how, etc., and in best practice in using email.

The only technical solution suggested, by several interviewees, was a request for an intelligent agent to filter incoming material, and/or scan for interesting things in a variety of sources.

The results showed interesting groupings among the users over the whole of their information-related behaviour, reflected in their thoughts on information overload. There were those who considered that overload was a result of improved provision, almost a good thing:

> 'Five years ago, we didn't have the information to be overloaded with, so we shouldn't complain.'
>
> 'We have more information, but also better tools to cope ... it was always possible to be overloaded.'

Those who took an entirely pragmatic view:

> 'I take what's needed and ignore the rest, nothing's changed.'
>
> 'I cope by not reading most things ... if it's important, I'll find out anyway.'
>
> 'It's a personal issue – people need, and can use, different amounts of information.'

And those for whom it was a real problem:

> 'We have more work than we can cope with, so naturally the same goes for the information.'
>
> 'When I see the pile of journals, I wish I were illiterate.'

**Is overload overestimated?**

It has been suggested above that information overload has been a phenomenon noted for a considerable time, and treated as an increasingly serious difficulty for information provision up to the present day. However, it should be noted that there has always been a countervailing view: that overload is an exaggerated difficulty; that, as it is sometimes expressed, one is overloaded to the extent that one wishes to be overloaded. This view was seen in the more pragmatic comments of the pharmaceutical researchers quoted above.

Rudd and Rudd[9], for example, argued that 'contrary to widely held belief', it was probable that library users rarely experienced information overload, that the belief that they did was due to confusion between the amount of information available (the 'information explosion') and its effect on users ('information overload'), and that true overload was likely to occur only under very specific circumstances. More recently, Exner[10] suggested that the average user has survived a period of 'drowning in information' and was now happily 'surfing the Internet'; the metaphors indicating that the overload problem has been avoided. Nicholas *et al.*[11] found in a survey of Internet use by British journalists that information overload was not a major issue. In a similar vein, some follow-up surveys, particularly among business users, to the Reuters study noted above have suggested that the overload problem is not now perceived as such a major problem.

Most controversially, Tildine[12] argues that information overload is frequently identified as a problem in popular culture, has not been properly documented through rigorous investigation, and is, in fact, a myth (in the strict sense of the term) of modern culture.

We do not find these viewpoints to be particularly convincing. It is true that information overload, like any other topic within information management, may be 'hyped-up', particularly by vendors with a solution to propose, and that the term may be over-used in general settings. It may be that, in some circumstances, such as use of Internet by particular groups, overload may be outweighed by other considerations. It also seems to be the case that some individuals will, simply on grounds of personality, decline to be overloaded. Nonetheless, the preponderance of research results, literature discussion, and professional experience lead us to believe that overload is a real, and continuing, issue.

## Causes of information overload

It may be argued that information overload is the natural and inevitable condition of the human species. Our senses, particularly the visual sense, are able to handle a huge input, and to identify significant patterns within it. The modern information environment, however, presents us with information in forms with which our senses, and prior experiences, are ill-equipped to deal. The causes of overload, in this sense, are multiple and complex; hence the difficulty in providing any single 'quick fix' solution.

This complexity is emphasised by the more thoughtful authors on the topic. Macusohn[13], for example, provides a cross disciplinary literature review of the notion of information overload for individuals working in organisations, and involving insights from neurology, psychology, management (organisation theory, organisational behaviour, and information management), sociology, and communication research. Simpson and Prusak[14], viewing information overload as a result of a failure in the business community to recognise the ways in which information processes add value to information, present a conceptual model of overload, based on the broad information attributes of truth, guidance, accessibility, scarcity and weight.

It is tempting, and usual, to assume that a major contributing factor, if not the only significant factor, in information overload is the TMI effect: 'too much information'. This is readily supported by statistics of the sort quoted by Schuman[15]:

- a weekly edition of the *New York Times* contains more information than the average person was likely to come across in a lifetime in seventeenth-century England;
- the English language of the late twentieth century contains about 50,000 words, five times more than in Shakespeare's lifetime;
- the collections of the large US research libraries doubled between 1876 and 1990;
- over one thousand books were published each day across the world during 1990.

Similarly, Infield[16] points to the proliferation of printed sources, web pages and email messages as a major cause of overload, while Biggs[17] identified 'the proliferation of available data and publications and ever-more-comprehensive and widespread automated means of access to them' as the cause of overload; a reminder that convenient access to information sources may be as much a cause of overload as the size of the information store.

It also seems clear that an increasing 'diversity' of information can also lead to overload; partly by virtue of a consequent increase in volume, information on a given topic may come from varying perspectives, but also because of an intellectual difficulty in fitting it within a cognitive framework appropriate for the use and the user. Iselin[18], for example, reports that increasing diversity of relevant information increased the overload effect for both structured and unstructured decision making. Diversity may occur both in the nature of the information itself, and in the format in which it appears, with a typical business user having to deal with paper, email, voicemail, Internet/intranet, and so on.

There is a general perception that new information and communication technologies (ICTs), aimed at providing rapid and convenient access to information, are themselves responsible for a high proportion of the overload effect. Certain kinds of technology are generally highlighted in this respect: electronic mail, and other messaging systems, and Internet/intranet.

Email generally appears high in users' perceptions of the causes of overload, both in formal surveys and in anecdotal experience: see, for example Palme, Karlgren and Pargman[19], and Losee[20]. Particular difficulties arise from such factors as overuse of 'blanket' email or needless 'cc-ing' of messages, from unfocussed information distribution within large newsgroups and mailing lists, and from 'spamming'.

Internet and intranet, though generally accepted as major information transfer tools, are also regarded as causes of overload, because of the sheer volume of information, often highly diverse, which they provide, coupled with multiple access routes: see, for example, Hamiltion[21], Ward[22], Goldstein and Lyon[23], Simmons[24], Black[25], and Basili[26]. Berghel[27] suggests that the most immediate cause of information overload on the web is caused by the web trying to fill the dual role of being both a private and a public information and communication medium; search engines, which should be able to deal with the problem, index 'more chaff than wheat', and are inherently ill-equipped to deal with the extreme diversity of web resources.

Information overload may also be caused as a result of the nature of the work being carried out by the user, with consequences for the nature and amount of information required. Interdisciplinary work in particular, increasingly common in many working environments, may be a clear cause of overload[28].

Another factor identified by some authors, including Biggs[17] and Basili[26], as a contribution to overload, is the greater proportion of information searching done by end-users rather than information professional intermediaries: so-called 'disintermediation'. The presumption is that users will not be so skilled as information specialists at rapidly identifying a core of valuable material, and hence may feel more overloaded.

## Solutions to information overload

As was noted above, information overload is a fundamental problem of the human condition, and has reoccurred throughout history. Much may be learned from past approaches, though the particular problems of today are novel. The methods used by organisations in the past were largely those of organisational structure, particularly hierarchical structures, and of standardisation and rationalisation. Innovations such as the tabulator, the standard-sized file and filing cabinet, standardised forms, the paperclip, and time zones, all arose as means of dealing with the 'crisis of control' in the nineteenth century, as speed of production and distribution outgrew the existing methods of control[29]. These methods in their turn have been outstripped by modern trends towards globalisation, and less tightly structured organisations, together with new technologies for the storage and communication of information.

Solutions to information overload, like its causes, are multi-faceted, and there is no single tool or technique that will correct the problem. The solutions which have been proposed or described appear to fall into two main categories: managerial and technical.

On the managerial side, the single most applicable and useful general concept in overcoming overload is that of 'control'. It was noted at the outset that loss of control over information is the single major symptom of overload, and restoring control is the major stem towards its remedy. Control may be exercised at both individual and organisational levels.

The term is used here, not in a coercive sense, but rather in its original meaning of comparison against a standard from the Latin, contrarotulare (to compare against the rolls); hence implying a purposive progress towards a predetermined goal, the opposite of the helplessness implied by being overloaded with information.

At individual levels, the skills required are neither complex nor technical. Many of them may be placed under the general heading of 'time management'. An obvious example is an explicit prioritisation of information seeking related to work goals and objectives, as opposed to a semi-random surfing among peripherally relevant material on the web, and an over-zealous reading of all incoming emails; see, for example, Infield[16], Rudd and Rudd[30] and Hopkins[31] make specific suggestions in this vein for the library setting. Setton[32] describes the individual means used by business executives to deal with overload, including dealing with a piece of paper as soon as it is touched (a key recommendation of many time management instructors), and deliberately seeking input from very different businesses. Many other recommendations – joining newsgroups and mailing lists very selectively, deleting many messages unread, only keeping material which it would be very difficult to find again – are, like most of time management, basic common sense. Alfred P. Sloan Junior's admonition to always consider what would be the worst that could happen if some information were thrown away is the epitome of this approach.

A concept generating increasing interest, and offering a partial solution to information overload at the individual level, is that of 'information literacy'[33], which has been termed 'the final key to the information society[34]. No single generally accepted definition exists, but in essence it refers to a set of information-handling skills going considerably beyond the simple ability to retrieve information, and usually including elements of evaluation, organisation and use. Perhaps the best succinct definition is that of Doyle[35]: 'information literacy is the ability to access, evaluate and use information from a variety of sources'.

The American Library Association's definition of the term includes six aspects:

- recognising a need for information;
- identifying what information would address a particular problem;
- finding the needed information;
- evaluating the information found;
- organising the information;
- using the information effectively to address the specific problem.

Gilster's[36] concept of 'digital literacy', and various other literacies, are variants on this theme; see Bawden[33] for a fuller description and comparison.

The relevance of this idea to the problems of information overload seems clear; it assists in the regaining of the control lost in the overload situation, by means of the clear recognition of what information is needed, and its evaluation and organisation. This involves a second concept which has had a good deal of discussion: critical thinking, in essence the establishment of mental frameworks to give a context to the mass of information which may be available on the subject at hand; for a fuller discussion see Bawden[33] and the papers in Oberman and Kimmage[37]. Promoting information literacy in individuals, by means of appropriate training, seems to be a powerful way of minimising the effects of information overload.

Personal information management training, on topics such as effective use of email folders, while an important part of this, is by no means the whole story; the more 'abstract' parts of information literacy should receive at least equal attention. Some training in classification and indexing will also be appropriate, for those who really need to keep large amounts of information.

At organisational levels, the remedies, at least at the first stages, are again likely to be simple and straightforward in nature. Promulgation of sensible rules for 'email etiquette' is a good example. This would cover such things as: not forwarding or copying material without careful thought; using 'blanket' mail should be used with great care – perhaps under central control; informative titles for all messages; attachments used sparingly, never for text only, and always with clear labels. Many organisations, particularly in North America, are adopting formal policies on email, restricting length of messages, rationing, or charging for, the sending of email, etc. Similar policies may be introduced for meetings, phone calls etc.

It is important that the preferences of individuals and groups are allowed for. People, and teams, differ in how they like to handle information, and hence how effective they are at different aspects of the issue. Enforced standardisation, in a well-meaning attempt to overcome overload, may well do more harm than good.

On the technical side, it is also, not surprisingly, suggested that, as new ICTs can be held responsible for a large part of information overload, so they provide something in the way of a solution. Many specific systems or functions have been suggested. These mainly fall into two classes: intelligent search agents and intelligent interfaces, for example Cawkell[38], Tudor[39], and Cooke and Kaupilla[40]; and systems for automatically ranking and filtering email and similar messages, for example Motiwalla[41], Palme, Karlgren and Pargman[19], and Losee[20]. A third general category is systems for customising retrieved information after it arrives. This may operate as a stand-alone facility, or may pick up where filtering systems leave off; see, for example, Berghel *et al.*[42]. Current examples of the customising approach are the refine/categorise features offered by Internet search engines such as AltaVista and Northern Light, which group retrieved sites into subject groupings by statistical analysis of terms; these should reduce the effort of examining large numbers of seemingly relevant web pages. Berghel[27], emphasising the importance of control in reducing overload on the Internet, recommends the development may be reduced by the use of personal software agents, information customisation tools, and brand names (for simpler identification of the required information), as partial solutions.

Although such systems are under active research and development, it is fair to say that none has yet demonstrated an ability to significantly reduce overload in an operational setting. It is, however, likely that such systems will be an important part of the solution for the future.

Finally, it is true to say that some commentators see a role for the information professional in helping users to combat overload, largely through exercise of 'traditional' information-handling skills; for example, creating catalogues of web resources[26], and providing advice of focused searching[31]. It is interesting to consider that the role of the information professional, which has hitherto been assumed to be that of identifying and accessing essentially all relevant information, may have to change to that of protecting users from information.

## References

1. Holmstrom, J.E. (1948), quoted by H East, 'Towards the schism.' *Journal of Information Science* 24(4), 1998, pp.217-275.

2. Torkelson N.R. 'SDI as a control of information overload.' *Proceedings of the 36th Annual Meeting of the American Society for Information Science, Los Angeles, 21-25 October, 1973.*

3. Reuters. *Dying for information.* (London: Reuters Business Information, 1996.)

4. Waddington, P. *Dying for information? A report on the effects of information overload in the UK and worldwide*. (London: British Library and Innovation Centre, 1997, pp.49-51.) (British Library Research and Innovation Report 78)

5. IIP *Nil by mouth*. (London: Investors in People UK, 1998.)

6. Wilson, P. 'Unused relevant information in research and development.' *Journal of the American Society for Information Science* 46(1), 1995, pp.45-51.

7. Hunt, R.E. and Newman, R.G. 'Medical knowledge overload: a disturbing trend for physicians.' *Health Care Management Review* 22(1), 1997, pp.70-75.

8. Bawden, D., Devon, T.K. and Sinclair, I.W. *Desktop information systems and services, a user survey in a pharmaceutical research organisation.* Submitted for publication 1999.

9. Rudd, M.J. and Rudd, J. 'The impact of the information explosion on library users: overload or opportunity.' *Journal of Academic Librarianship* 12(5), 1986A, pp.304-306.

10. Exner, F. 'From drowning to surfing: a slogan's significance.' *Journal of the American Society for Information Science* 49(1), 1998, pp.92-93.

11. Nicholas, D. *et al.* 'The Internet: it's early days, but there are some surprises.' Aslib Proceedings 49(8), 1997, pp.214-216.

12. Tildine, T.J. 'The mythology of information overload.' *Library Trends* 47(3), 1999, pp.485-506.

13. Macusohn, L.M. 'The information explosion in organisations: a review of individuals' information (over)load.' *Svensk Biblioteks Forskning* (3-4), 1995, pp.25-41.

14. Simpson C.W. and Prusak, L. 'Troubles with information overload: moving from quantity to quality in information provision.' *International Journal of Information Management* 15(6), 1995, pp.413-425.

15. Schuman, P.G. *Introduction to Information literacy: learning how to learn*. Varlejs, J. (ed.) (Mcfarland, Jefferson NC, 1991, p.4.)

16. Infield, N. 'Too much information?' *Library Manager* 14(25), 1996.

17. Biggs, M. 'Information overload and information seekers: what we know about them, what to do about them.' *Reference Librarian* 25/26, 1989, pp.411-429.

18. Iselin, E. 'The impact of information diversity on information overload effects in unstructured managerial decision making.' *Journal of Information Science* 15(3), 1989, pp.163-173.

19. Palme, J. Karlgren, J. and Pargman, D. 'Issues when designing filters in messaging systems.' *Computer Communications* 19(2), 1996, pp.95-101.

20. Losee, R.M. 'Minimizing information overload: the ranking of electronic messages.' *Journal of Information Science* 15(3), 1989, pp.179-189.

21. Hamilton, F. 'Internet daze.' *Information Management Report*, August 1997, pp.1-4.

22. Ward, C. 'Coping with information overload.' *.net* 38, 1997, pp.70-74.

23. Goldstein, M. and Lyon, J. 'Seize the hype.' *Information World Review* 128, 1997, pp.24-25.

24. Simmons, S. 'The Internet for NHS managers: endless potential or information overload?' *Managing Information* 4(3), 1997, pp.40-44.

25. Black, G. Intranets change the rules, Information Management Report, May 1997, pp.5-17.

26. Basili, C. 'Verso la societa dell'informazione.' [Towards the information society.] *Biblioteche Oggi* 16(6), 1998, pp.50-53.

27. Berghel, H. 'Cyberspace 2000: dealing with information overload.' *Communications of the ACM* 40(2), 1997, pp.19-24.

28. Wilson, P. 'Interdisciplinary research and information overload.' *Library Trends* 45(2), 1996, pp.192- 203.

29. Beniger. J.R. *The control revolution: technological and economic origins of the information society*. (Cambridge, Mass: Harvard University Press, 1986.)

30. Rudd, J. and Rudd, M.J. 'Coping with information load: user strategies and implications for librarians.' *College & Research Libraries* 47(4), 1986B, pp.315-322.

31. Hopkins, R.L. 'Countering information overload: the role of the librarian.' *Reference Librarian* 49/50, 1995, pp.305-333.

32. Setton, D. 'Information overload.' *Forbes* 160(6), 1997, pp.18-20.

33. Bawden, D. *Information literacy, digital literacy and other literacies: a review of concepts and models*. Submited for publication 1999.

34. Oxbrow, N. 'Information literacy – the final key to an information society.' *Electronic Library* 16(6), 1998, pp.359-360.

35. Doyle, C.S. *Information literacy in an information society: a concept for the information age*. (Syracuse, NY: ERIC Clearinghouse, ED 372763, 1994.)

36. Gilster, P. *Digital literacy*. (New York, NY: Wiley, 1997.)

37. Oberman, C. and Kimmage, D. (eds.) *Russian American seminar on critical thinking and the library. Occasional papers 200/201*. (Urbana-Champaign, IL: Graduate School of Library and Information Science, University of Illinois, 1995.)

38. Cawkell T. 'Could a search agent solve the information overload problem?' *IT Link* 10(7), 1997, pp.3-5.

39. Tudor, J.D. 'The new alchemy using droids & agents to treat information overload.' *Online* 21(6), 50-58.

40. Cooke, S. and Kaupilla, A. 'Tools for the Internet: channelling the information tidal wave.' *Electronic Library* 14(2), 1996, pp.129-132.

41. Motiwalla, L.F. 'An intelligent agent for prioritizing e-mail messages.' *Information Resources Management Journal* 8(2), 1995, pp.16-24.

42. Berghel, H. *et al.* 'Cyberbrowsing: information customization on the web.' *Journal of the American Society for Information Science* 50(6), 1999, pp.505-513.

# Derivation of interaction environment and information object properties for enhanced integrated access and value-adding to electronic documents

*Chern Li Liew and Schubert Foo*

*Division of Information Studies, School of Applied Science, Nanyang Technological University, Nanyang Avenue, Singapore 639798*

*Email: PH8598467@ntu.edu.sg; assfoo@ntu.edu.sg*

**Abstract** *The recent unprecedented growth of information in digital form advocates the need for better ways to interact with this vast amount of richly-conveyed information. This paper reviews and analyses the design considerations of an advanced information environment for users to interact in novel manners with online documents such as electronic journals. The interaction environment is also intended to support value-adding of electronic documents. The design is inspired by analyses of user requirements in interacting with information, by advances in the electronic information world, and by innovations in human-computer interaction. These led to the derivation and proposal of a set of properties in both the interaction environment and the information objects within the environment. The environment is a visually-rich, interactive information environment based on user-controlled malleability and integration with various interaction tools. The information objects within the environment are fractionally structured, contextualized and explicable, queriable and navigable at multiple levels of granularity, and associated with layers of additional information and metadata.*

## Introduction

Information in digital form is fast replacing its traditionally printed counterparts as the result of increased computing power, information processing tools and networked connectivity. Full text journals, electronic newspapers, encyclopaedias and other reference sources are becoming commonplace. Research has increased to examine and exploit the potential of using advanced technologies to provide enhanced information. Examples of such activity include the web journal projects[1,2] as part of the e-Lib Programme[3] and work by Williams and Nicholas in the provision of value-added news information[4]. Digital information is more abstract, dynamic and malleable in comparison to its printed form. Hence, an understanding of how to enhance its value and its interaction becomes an important prerequisite for the new electronic information age that is rapidly dawning upon us. The user interface of future electronic documents (e-documents) should be thought of as a complete workspace in which a variety of information processing or manipulation tools are available for users to flexibly interact with richly-conveyed information. Researchers are exploring various approaches and the utility of various tools to improve and enhance interfaces to e-documents. The research of which this paper is a preliminary report joins this effort.

This paper examines how the electronic information environment is likely to change in order to enhance the user-information interaction and to support the value-adding of e-documents. This requires the understanding of the various issues involved in such interaction and an exploration of the design space of the user interface. Through a review and analysis of related work, a set of basic desired properties for both the interaction environments and the information objects within the environments is derived.

## Related work

Understanding the electronic information environment requires an understanding of the information system, the user needs and the technologies that are available to support the user interaction with information. Hence, issues in the design of interactions with value-added e-documents arise from three main sources:

- increasing knowledge of how users interact with (seek and use) information, and the tasks they perform to achieve their goals;
- the ongoing development of the information world, leading to a new paradigm of electronic information; and
- advances in user interface technologies that have provided the basis for much more elaborate interactions with the underlying information world.

As such, the review and analysis of related work is presented along these three inter-related areas.

### *User information interaction and needs analyses*

Adopting a user-centred approach to information system design requires a thorough understanding of user needs and their real interactions with information. Attention to users of digital information systems and their needs from a communications perspective have been well presented as various models in literature. For the purpose of this research, a subset of these proposed models that holds more direct implications on the design of interactive interfaces for end-users is presented and reviewed. Alongside the review, the significance of these findings on the design of the information environments for future e-documents is identified.

Belkin and colleagues[5] propose that information systems need to capture users' Anomalous State of Knowledge (ASK). When users (information seekers) approach an information system to resolve an anomaly in their state of knowledge, they are not aware of exactly what the anomaly is and the type of information needed to resolve that anomaly. Hence, users are unable to specify precisely what they need at all times. By observing information seeking in a variety of settings, a set of information seeking strategies that include the goal of interaction (learn/select), method of interaction (scan/search), mode of retrieval (recognize/specify) and type of resource (information/meta-information) is proposed[6,7]. Accordingly, users must go through a process of clarification to articulate a search request, with the obvious implication that electronic systems should support iterative and interactive dialogues with users. The system must therefore allow users to scan, recognize, learn and select, amongst others.

Bates[8] proposes a berrypicking model to illustrate the evolving search process in using online and other information systems. Users engage themselves in an active and iterative process with the information systems using the traditional search-retrieval paradigm as well as the exploratory paradigm to locate and retrieve the relevant information. Thus, users might consult a variety of sources and pick relevant items one at a time during the search-retrieval process. Bates focuses on and describes the browsing concept and subsequently develops six strategies associated with this concept; namely, footnote chasing ('backward chaining'), citation searching ('forward chaining'), area scanning, subject searching in bibliographies and A&I (abstracting and indexing) services, and author searching. Every new piece of information the searchers encounter has the potential to provide them with new ideas and directions to follow. This berrypicking model depicts the need for interfaces to support these strategic, opportunistic interactions with information.

Another researcher, Ingwersen[9] proposes a polyrepresentation approach in information retrieval where user's cognitive space not only represents user's current (often topical) information need, but also more importantly, the underlying problem space, actual work task or interest, and the dominant work domains. This approach implies that different methods of representation and a variety of search and retrieval techniques of different functional origin should be available simultaneously to each semantic entity in an electronic system. However, user interactive process, such as that of feedback, are only implied processes within this model. The challenge for a more interactive model is to elaborate and incorporate feedback and user judgments related to, for instance, query reformulation, visualisation and navigation.

Cool and fellow researchers[10] discuss one particular type of interaction with text – that with the goal of evaluating usefulness. This involves looking first at characteristics of the text such as author or title, to determine the topical relevance of a specific document. Meanwhile, Belkin[11] examines another type of interaction – that with the goal of discovery, whereby users scan the whole texts widely or part of the text to determine its relevance. These two studies indicate two different requirements on text representation and structuring in support of effective interaction with information. The formal study shows that characteristics of secondary texts will need to be represented and related to one another, and other attributes of texts, in forms that can be used in ways other than for bibliographic specification, or as retrieval leads. The latter study indicates that facilities for dynamic, user-defined and large scale patterns in texts are required, and patterns should be capable of being related to other sections of the texts through dynamically established clusters or links.

Pirolli and Card's[12] information foraging model attempts to explain how various system design criteria could affect the balance between the value of information gained and the user's cost of performing various activities under the different designs. Pirolli[13] later presents a more elaborate description of the technique and its application to exploring the design space for a browser for very large text repositories. This approach permits the exploration of 'what-if' design scenarios and the exploration of changes in ideal user strategies across system and task conditions.

Saracevic[14] recommends a stratified model of interactive information retrieval. The actors in the model comprise of the users, computers and intermediaries. Interactions between the actors are believed to be sequences of processes advancing at different interconnected levels or strata. At the surface level, dialogues between users and systems are captured, where users express their needs with the computer and the computer responds with results. At the cognitive level, users interpret information provided by the system. At the situational level, the context or situation in which users approach the system motivates the interaction. The different strata are closely interrelated and are extremely dynamic with each move by the user or system influencing other moves or shifts within interactions, such as information problem, strategies, feedback or uncertainty. This leads to major implications for interface design whereby a user's relevance judgments and search episodes (i.e. interactions with systems) must be visually analysed and presented to the user in a visible manner.

Apart from these generic models, a subset of relevant studies that concentrates on end-users of electronic systems exists. Dillon[15] acknowledges the differences between using a printed document and its electronic counterpart. He concludes that future e-documents will need to have more powerful navigation and manipulation facilities to support the task users need to perform, to effectively use these e-documents. Studies of end-users are sometimes part of an overall development effort for a specific electronic system[16, 17, 18]. Generalized results from these studies essentially elucidate the high acceptance of interactive browsing and navigation by typical end-users as the main strategy employed in using e-documents. Among the features identified as important support that a system must provide are:

- to help in the user question-negotiating process;
- allowing the user to pursue multiple approaches of inquiry by entering fragments of the question;
- allowing the user to explore multiple navigation options; and
- to capture partial results – modelling the user's current interest and perhaps, suggesting the next navigation choices according to the nearness of this interest.

The way information is presented to the user is also a crucial issue. To deal with the common information overload and screen cluttering problem, Kreitzberg[19] proposes a 'just enough up front, details on demand' paradigm in designing e-documents. He suggests providing just enough information initially to ensure comprehension and presentation of information in a logical sequence so that the information which has come before provides context for that which follows, and imparting details to users on demand. Kreitzberg's paradigm stresses the importance for information units to maintain conceptual integrity and comprehensive coverage, by treating each unit as an initial presentation, which links to pre-requisite concepts required for understanding, and avoids cluttering the presentation with unnecessary details.

The Open Journal Project[20], which is concerned with user interaction with electronic journals (e-journals), has received a number of interesting reviews from its access demonstration of citation linking in the *Open Journal of Cognitive Science* in April and May 1998. Users generally favour the link structure and demand for more links to trace ideas and in serendipity searches. They also call for better orientation and additional navigation aids to keep track of where they are in the clusters of links. Dillon, Richardson and McKnight[21] note that users generally first select and scan sections of a journal article before judging the relevance and probable quality of the document. While scanning through the article, users often prefer to maintain an overview of the whole article while exploring the different sections (for example, introduction, section headings, diagrams/tables and conclusion) before deciding to read on, or to reject the article. The implication for the design of e-journals is that the system must support this exploration while providing an overview, and offer features to help users to decide the relevance of the article, or to perform other tasks. Such features can include, for instance, display explicitly a list of other papers citing the current article, word or phrase dependencies and occurrences within the article, and display a list of other works published by the article's author(s). All these features add value to the information while maintaining an overview of the context. As researchers begin to move towards collaborative work, support of annotations both for personal and shared use may also be desirable. If users of e-journals are to find required and desired information more quickly and better-placed in context, future e-journals will have to strive for completeness of information presentation and be supported by advanced technologies in the organisation and presentation of the information.

In addition to the major user-information interaction models and understanding the user needs and requirements, it is also necessary to examine the user interface requirements. Between the analytical and browsing strategies of searching, this research emphasizes the more heuristic browsing strategy, particularly, in improving the user interaction with information units within e-documents. Hence, tasks such as hierarchical traversal, cluster-based traversal, dynamic links traversal, panning and zooming for context and details, and accessing a variety of information sources in a seamless manner becomes more significant in this context. Shneiderman[22] proposes a framework of seven tasks that form the basic interface features that the next generation of information systems interface should support. We trust the framework to encompass the major functionality required to support effective user-information interaction in this research. These seven tasks are: overview, zooming, filtering, details-on-demand, relate, history and extract.

An overview provides a good starting point for the user to 'scan and select' and to provide the overall information context. Smooth zooming helps users to maintain their sense of position (context) while exploring the various links or sections within an e-document. Filtering is used to weed out uninteresting or unnecessary details in a document or collection to help users focus on relevant details. Details-on-demand provides the option to allow the user to take a closer look at details (e.g. full-text that stands behind any representation or attached files) of one or more objects in the field of view. Relate makes explicit relationships between objects in a display. It may also refer to representing relationships between data in multiple associated windows. A history is useful to maintain a list of previous actions in order to support place-keeping, undo and for progressive refinement. Finally, extract permits users to extract an item or a set of items of interest from the information context for further exploration.

### *Developments in the electronic information world*

A second area of consideration is the advancement of the electronic information world, where the view of the possibilities for new dynamic information entities and more advanced interactions with electronic information has been expanded. Information has become ever more digital, enabling, for instance, more uniform interactions where all format can be stored and sent equivalently. This has also made it possible for various parties (e.g. author, publisher/information broker, other third parties) to package metadata along with primary content. The importance of metadata is growing in various niches and applications can utilise metadata to integrate with one another, to exhibit the relationships with one another, and also to manifest the larger context in which they are being used. It is possible, indeed, to apply more than one action to an information object, something that currently lies outside the standard World Wide Web paradigm.

Information structure is also becoming valuably rich – information is making accessible its structure as, for instance, text and metadata formats move to SGML[23]. Newer SGML-based metadata formats (e.g. TEI Headers) can be distinguished from the simple structured generic formats (e.g. Dublin Core) by fullness (delving into more details), structure (contain richer structuring schemes) and by specialism (they may be more specific to the relevant domain). They can be used for location and discovery, and in documenting information objects and/or collections of objects. They allow for some level of evaluation and analysis of content, and navigation around aggregations of objects. They are also expressive enough to capture various relationships at different levels of aggregation. However, to make information more apparent to more general resource discovery tools, export of data in some simpler formats such as the Dublin Core[23] may be needed.

The transition from the physical delivery of printed documents to electronic documents in the form of bits and bytes over the network has made the granularity of information packaging variable. This consequently results in the need to interact with information at many different scales. The requirement to deal with a massive electronic world and the emergence of distinct structure has led to fractional, modular[24], layers[25,26] of information organisation. Indeed, in the electronic environment, from super-digital library collections down to individual documents, down to pages, paragraphs and diagrams, even to the level of words, can now be teased apart and be manipulated to get more value out of the different levels of information objects. As we move from an environment where information is transformed from linear to multidimensional form, information objects no longer need to fit into any physical unit but comprising logical components of information, probably built-up of units each of which could stand independently on their own. In such a case, each of these information objects will need to have sufficient descriptive information to enable it to be used in various contexts. SGML is an example that provides flexible structuring capabilities to describe these objects.

These considerations of new forms of information presentation and structuring have important implications for the design of a new interface and interaction environment with e-documents. Interfaces are starting to move beyond mimicking mechanisms of earlier media and traditional interface metaphor to more thoroughly exploit computationally-based mechanisms, such as that proposed by Mohnkern[27], to deal with information that is much more dynamic and reactive to user needs and preferences.

### *Developments in UI technology*

Just as the electronic information world has advanced, so has the technologies that form the basis for information worlds. Advances in user interface technology have recently been reviewed by Myers[28] and Shackel[29]. The past 15 years of advances have started to provide the foundation for much more elaborate interactions with the underlying information world – with collections of documents and with individual documents through graphical or visual interfaces. Visual displays of information are fast gaining acceptance and recognition[30]. This is because visual displays are more easily understood, providing an operating environment where an explicit visual and functional context for the computer user's actions is formed[31]. It is also important to understand at the onset that the user interface is not only a screen design, but a method of interacting with the application and its data. Interaction techniques are hence, an integrated part of visualisation. User interfaces to e-documents which offer greater sophistication and technological mastery in manipulating diverse forms of information will be in great demand. Young generations of computer users are becoming 'digital sorcerers'[32] – handling new computer games, software applications and personal digital assistants with ease. Apart from the combined efforts of researchers and user-oriented designers, this phenomenon is also likely to be attributed to the intimacy derived from simply living with advanced technology. It is almost assured that future generations of e-document users, with their grasp of digital manipulation, will demand new forms of interaction and redefine current notions of creative expression.

Two of the more recent and more relevant developments in the user interface technology, that will form the basis for more powerful interaction with future e-documents, are the construction of direct manipulation (DM) interfaces in the 1980s, and especially, the ensuing emergence of information visualisation (IV) techniques in the 1990s. DM makes interaction with electronic information more intuitive. Its applications respond interactively to the user's input actions. Compared to traditional information access tools that are limited in their ability to display results, and often require users to remember obscure command vocabularies and complicated syntax, DM allows users to visualise and directly manipulate objects on computer screens. Recent works in the 1990s have carried the user's desire for manipulating real objects still further. Current works on the *FilmFinder*[33] and the *HomeFinder*[34] for example, have shown that users are able to locate information faster than with natural language queries. User comprehension and satisfaction have also proven to be higher for these interfaces.

The success of DM interfaces is a testament to the power of using computers in a more visual, or graphical manner[22]. Existing IV systems that allows users to zoom, filter and examine details-on-demand are built on a similar principle. Users not only desire to find information quickly in the electronic environment, they also want comprehensible and commandable interfaces to e-documents that will provide them with a feeling of accomplishment and empowerment. Users require a system where the focus is on DM of objects presented, using a virtual reality type of paradigm to promote real-time feedback, and be given intuitive control over where and how the exploration is occurring. Such a system should also not try to display all the potentially very large and dense, multidimensional data at once. Users must instead, be allowed to interactively select fragments of interest and explore selected information. IV is required to complement and enhance current e-document interfaces. These techniques will enable users to conduct their searching and browsing tasks more efficiently and creatively.

There are parallels between IV and scientific visualisation, in terms of purpose and method – the purpose being that of extracting prominent dimensions from multidimensional data, and the method being that of using advanced graphical and animation techniques to present data. Spence[35] points out that visualisation is something a human engages in and that its potential value lies in gaining insights and understanding. He describes the function of IV as the presentation and representation, and rearrangement and dynamic exploration of information, whereby humans interact with information to gain dimensionality and perception in order to comprehend.

In the application of IV, consideration needs to be given to the fact that the interesting information is mostly abstract (e.g. information structures and relationships) so that intuitive comprehensible visual metaphors have to be developed. Also, in contrast to scientific visualisation, where applications are focused on trained scientists and engineers, interfaces created to manipulate e-documents may be widely employed amongst a diverse and potentially non-technical population that varies in terms of their levels of education, background, experiences, capabilities, information needs, and the tasks they need to perform. Hence, one of the main goals of the application of IV should be to enable these disparate user groups to use visual presentations that will be tailored to their specific needs. The tools and the interface must also be intuitively easy for each of these groups to handle.

In e-documents such as e-journals that are laden with data and information, appropriate utilisation of IV has the potential to assist users in identifying, exploring, discovering and developing understanding of possible complex situations in interacting with e-documents. IV also supports the value-adding of e-documents. Value-adding comes from the ability to enable manipulation and analysis of both the presentation and attributes of data to produce possibly new information that might not otherwise be obvious or even discernible.

From this analysis of the existing IV techniques and an earlier expansion[36] of previous

*Table 1. Taxonomy of information visualisation techniques*

| **IV Taxonomy** | **Level of Applications & Content** | **Main Tasks Examples of Exploration** | **Examples of Potential IV Techniques (in no particular order)** |
|---|---|---|---|
| **InfoSphere Visualiser** | Infosphere – information outside user's work environment (e.g., information on WWW or in a digital library) | For an overview – to obtain a sense of overall composition and layout of the information space.<br>To help locate information (e.g. within a networked collection). The visualisation can take the form of a virtual space that contains the information collections. | • *Pad++*[i]<br>• *Galaxy of News*[ii]<br>• *Butterfly Citation Browser*[iii]<br>• *WebTOC*[iv]<br>• *Elastic Windows*[v] |
| **Information Structure Visualiser** | Information Workspace – information the user is interacting with as part of an activity. | For an overview.<br>To view the large structure of the user's information workspace in a complete, coordinated manner. | • *Perspective Wall*[vi]<br>• *Document Lens*[vii]<br>• *Cam/Cone Trees*[viii]<br>• *Fisheye View*[ix]<br>• *Hyperbolic Tree Browser*[x]<br>• *WebBook & Web Forager*[xi] |
| **Visual Information Probes** | Visual knowledge tools – a set of information objects. | Used for zooming, filtering and details-on-demand.<br>The concept of lens, filters and other tools that provide users with the ability to examine a particular part of the information world in multiple ways and to help detect trends and patterns. | • *Magic Lens*[xii]<br>• *Movable Filters*[xiii]<br>• *Table Lens*[xiv]<br>• *Influence Explorer*[xv]<br>• *LensBar*[xvi]<br>• *TileBars*[xvii]<br>• *SeeSoft*[xviii]<br>• *Fractal Views*[xix]<br>• *Visage*[xx] |
| **Enhanced-information-objects Visualiser** | Visual objects – one or more data sets integrated together for convenience of search and navigation. | Used for making relationships explicit between objects, history tracking and extracting an item (and sets of items). | • *Envision*[xxi]<br>• *LifeLines*[xxii] |

works done on IV classifications[37,38], we propose that IV techniques be taxonomized into four main categories as shown in Table 1. This taxonomy takes into consideration the level of application[39] and the seven tasks framework proposed by Shneiderman[22]. In the Table, the main categories that include the InfoSphere visualiser, information structure visualiser, visual information probes and enhanced-information objects visualiser are accompanied by descriptions of the level of application and content, main tasks and examples of existing IV techniques.

### *Summary of related work*

From the review of related work, some pertinent observations become obvious. An innovative information workspace with new interaction approaches is desired for users to interact with electronic data in novel, multiple complementary ways. This becomes ever more important as our lives are increasingly being conducted on a base of information technology suffused with computationally cross-accessible information. Users must not be seen to be passive recipients, but instead, active composers and manipulators of information. Users must be also allowed to re-organise the information where necessary to create new information that meets their needs. Through the application of these novel tools, value-adding to e-documents may also be achieved.

From the examination of the existing IV techniques, it is apparent that no single IV technique is sufficient for the kind of interaction and exploration desired in the information environment. A suite of integrated methods will render a better environment, with various techniques being used in different contexts. Just as information seekers apply various searching and browsing strategies according to their tasks, different mechanisms will be appropriate for different situations.

## Design desideratum for the information environment

Based on the preceding reviews and analyses, a set of desired properties to permit enhanced interaction and value-adding to e-documents is identified and proposed. These are divided into interaction environment properties and information object properties.

### *Interaction environment properties*

Interaction environment properties are those properties that are closely associated with the user interaction with the information environment. The e-document environment should therefore be endowed with the following desirable properties.

1. *Integration of workspace* – the information environment should integrate both the user's personal workspace and the broader information world to allow a tighter connection between the user's own larger task that is normally centred in the private workspace, and their information-seeking sub-tasks that are often focused in the broader information world. This kind of integration facilitates the launch of queries based on current work and, therefore, enables a uniform paradigm for accessing both one's own workspace and the information world at large. Integration also simplifies the task of consolidating intermediate and final information seeking results into one's initial task.
2. *User-controlled malleability* – the environment must support DM to enable users natural, interactive access to the various information objects and tools. Users should, for instance, be given the ability to re-arrange or zoom in for more details on selected objects when required and desired. Users can explore, navigate and perform queries almost simultaneously in such an environment. All the IV techniques we have examined allow users to perform such direct manipulation on objects.
3. *Visual sovereignty* – the electronic environment should also provide the user with visual representation to the various conceptual items of the user's information-seeking task (e.g. a collection of information objects related to the task, or the user's search and browse history records). This is so that the user can be made aware of these details and their status, and be able to act upon them according to their

needs. Providing such a 'total view' while searching gives the user the satisfaction of being able to perceive the existence of related objects.

4. *Multiple representation/views* – multiple representation of information objects and their organisation should also be a part of the environment, so as to support a variety of task and information-seeking perspectives. The information seeker should be able to manipulate various aspects of the representation and view them directly, and they should also be able to make a choice of the view of interest to them, according to their needs.
5. *Integration with appropriate tools* – the environment should also provide tools for basic common tasks users perform in information-seeking and exploration, which include querying for information, the more heuristic navigation through information space, organizing and authoring information. Querying, for instance, requires an interface that will allow users to specify the query. Browsing or navigation requires information to be organized and corresponding visual clues about where items are; users should be allowed to weave through the information world to locate what is available, and to seek specific targets by, for instance, pan and zoom mechanisms. The environment should enable intuitive and easy means for users to move back and forth between the more informal cursory process and the more detailed, focused efforts in information-seeking.

### Information object properties

Correspondingly, the information objects in the environment should also be endowed with a number of characteristics to support a variety of information-seeking and gathering tasks. These properties should hold for all the information objects in the interface, including objects deposited in remote collections (i.e. those in the wider infosphere) and objects (e.g. annotations) created by users themselves.

1. *Structured* – information objects should have structures both internal and external to themselves. They, for instance, can enclose, or be contained in sets, collections or ordered lists, trees, tables, natural language sentence structures, sections or chapters. Thus, these objects should make visually explicit their internal structures to help users in the information searching process. They should also state explicitly where they fit into various larger structures.
2. *Fractional* – objects should allow manipulation and interaction at all granularities in a coherent manner. Users for instance, should be able to view, navigate, search and manipulate explicitly at the level of collections, sub-collections, documents, sections, paragraphs, sentences or words. It is interesting to note that in an electronic environment, seeking within a document now often includes across-document seeking as well. The distinction between one single document and a collection of documents has been blurred. This fractional aspect of structure not only reflects the new perception of the reality of the new information sphere, but also of information seekers' activities. Analyses have shown how users often engage in moving down levels of granularity as they home in on target clusters of information, or up as they backtrack to seek elsewhere.
3. *Queriable and navigable* – all objects in the infosphere at any granularity should also be queriable and navigable. First, there is the 'every item is a collection' effect, whereby a query can be directed at any object at any granularity. Users can, for instance, specify a search to a document-object, paragraph-object, table-object or word-object. Taking into consideration the internal and external structure discussed earlier, all objects then must provide the equivalent of an index of their internal contents. The objects also must impart index-relevant information to the external objects that contain them. Additionally, 'any item can be a query', whereby not only words, but phrases, paragraphs or the whole document can be used as queries against other objects

in the environment. Users should also be able to browse and navigate their way towards items of interest. To support this, objects must have visible structure and come with ways to choose between, or move around, within their various subfraction.

4. *Explicable* – as in general, information-seeking and knowledge work is a temporarily prolonged process, with information seekers often moving back to previous work done and building on it, all objects should ideally come with representations of their history – where they originate from and what operations have been performed on them. The historical sequence of the objects generated should be presented in a visualizable, malleable history mechanism.
5. *Contextualized* – information objects should make various aspects of their context such as relationships between various contents (both established by system and the user) available to the user, to enable better understanding of both the objects and the environment. This is important to support the context-driven evolution of a user's interests.
6. *Affiliation with metadata and additional layers of information* – evaluation, reviews, commentaries and criticisms are all value-adding properties that provide information seekers with additional information for decision-making. Annotations on all objects at any granularity, both for personal and shared use, should also be supported. This will become even more important when people no longer work in isolation, and where technology is used for collaborative work. Annotations are also value-adding properties. It is evident that all the information objects should also be associated with a large prescription of metadata to implement many of the discussed properties.

Metadata can also be used to add value to objects. Examples of meta-level value-adding include:

- information customisation (e.g. non-prescriptive, non-linear document traversal, variable links and integration of sources);
- searching, indexing and database technologies (e.g. provide dynamic, real-time analysis for document clusters, with innovative mining and clustering topologies for display of results).

In a commercial sense, some of the value-added elements maybe packaged (tailored) and distributed (and possibly charged for) separately (i.e. as add-ons or as part of a customized document) to various customers with disparate needs.

## Conclusion

This paper identifies and proposes a set of interaction environment and information object properties to support the enhanced interaction and value-adding of e-documents, such as e-journals. These are inspired by background reviews and analyses of user-information interaction requirements, of the evolving forms of information structuring and presentation in the electronic information world, and of novel interactive user interface technologies. The research advocates that these properties form the foundation features for a range of new interfaces that are designed to provide an enhanced level of user-information interaction and to support e-document value-adding.

Such a prototype interface is currently under design and study at the Division of Information Studies, Nanyang Technological University, Singapore. The interface operates in a visually-rich interactive environment based on the derived set of properties and the ensuing interaction paradigm smoothly merges visualisation, querying and navigation of information. It forms part of a concerted effort in the research and development of a Digital Library for Creative Learning and Research (DLCLR)[40] environment in the School of Applied Science.

A follow-up paper detailing the interaction environment together with a presentation and discussion of a set of different scenarios of how such an environment is likely to be used for interacting with e-journals is currently under preparation. These scenarios are used as a 'proof of concept' to illustrate the ability of the interface to provide enhanced user interaction and value-adding. It is also aimed at demonstrating how digital information and user interaction is likely to be perceived and used in the future. At the same time, the scenarios form the basis for the design of the usability experiments that accompanies the interface design process that will be carried out in the next phase of this research.

## References

1. JISC. 'Open Journal Project.' Available at: *http://journals.ecs.soton.ac.uk/* (visited April 1999).

2. JISC. 'SuperJournal Project.' Available at: *http://www.superjournal.ac.uk/sj/* (visited April 1999).

3. 'e-Lib Electronic Libraries Programme.' Available at *http://www.ukoln.ac.uk/services/elib/* (visited April 1999).

4. Williams, P. and Nicholas, D. 'The Internet, a regional newspaper and the provision of 'value added' information.' *Aslib Proceedings* 50(2), October 1998, pp.255-263.

5. Belkin, N.J. *et al.* 'ASK for information retrieval: part I. Background and theory.' *Journal of Documentation* 38(2), 1982, pp.61-71.

6. Belkin, N.J. *et al.* 'Cases, scripts and information seeking strategies: on the design of interactive information retrieval systems.' *Expert Systems with Applications* 9, 1995, pp.379-395.

7. Belkin, N.J. 'Interaction with texts: information retrieval as information seeking behavior.' *Information Retrieval*, 1993, pp.55-66.

8. Bates, M. 'The design of browsing and berrypicking techniques for the online search interface.' *Online Review* 13(5), 1989, pp.407-424.

9. Ingwersen, P. 'Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory.' *Journal of Documentation* 52, 1996, pp.43-50.

10. Cool, C. *et al.* 'Characteristics of texts affecting relevance judgments.' In: Williams, M.E. (ed.) *Proceedings of the Fourteenth National Online Meeting*, 1993, pp.77-84.

11. Belkin, N.J. 'Design principles for electronic textual resources: investigating users and uses of scholarly information.' *Current Issues in Computational Linguistics*, 1994, pp.1-18.

12. Pirolli, P. and Card, S.K. 'Information foraging in information access environment.' *Proceedings ACM Computer-Human Interaction – Annual Conference on Human Factors in Software*, 1995, pp.51-58.

13. Pirolli, P. 'Exploring browser design trade-offs using a dynamic model of optimal information foraging.' *Proceedings ACM Computer-Human Interaction*, 1998, pp.33-40.

14. Saracevic, T. 'The stratified model of information retrieval interaction: extension and application.' *Proceedings Annual Meeting of the American Society for Information Science* 34, 1997, pp.313-327.

15. Dillon, A. *Designing usable electronic text – ergonomic aspects of human information usage.* (London: Taylor & Francis, 1994.)

16. Marchionini, G. and Crane, G. 'Evaluating hypermedia and learning: methods and results from the Perseus project.' *ACM Transactions on Information Systems* 12(1), 1994, pp.5-34.

17. Borgman, C.L. 'Why are online catalogs still hard to use?' *Journal of American Society for Information Science* 47(7), 1996, pp.493-503.

18. Hirashima, T. *et al.* 'Context-sensitive filtering for browsing in hypertext.' *Proceedings Intelligent User Interfaces,* 1998, pp.119-126.

19. Kreitzberg, C.B. 'Designing the electronic book: human psychology and information structures for hypermedia.' In: Salvendy, G. and Smith, M.J. (ed.) *Designing and using human-computer interfaces and knowledge-based systems.* (Amsterdam: Elsevier Science, 1989, pp.457-464.)

20. Hitchcock, S. *et al.* 'Linking electronic journals – lessons from the Open Journal Project'. *D-Lib Magazine*, December 1998. Available at: *http://www.dlib.org/dlib/december98/12hitchcock.html* (visited April 1999).

21. Dillon, A. *et al.* 'Human factors of journal usage and design of electronic texts.' *Interacting with computers* 1(2), 1989, pp.183-189.

22. Shneiderman, B. '*Designing the user interface: strategies for effective human-computer interaction, 3rd edition.*' (Reading, Massachusetts: Addison-Wesley, 1998.)

23. Dempsey, L. and Heery, R. 'Metadata: a current view of practice and issues.' *Journal of Documentation* 54(2), March 1998, pp.145-172.

24. Kircz, J.G. 'Modularity: the next form of scientific information presentation?' *Journal of Documentation* 54(2), March 1998, pp.210-235.

25. Phelps, T.A. and Wilensky, R. 'Toward active, extensible, networked documents: multivalent architecture and applications.' *Proceedings First ACM International Conference on Digital Libraries*, 1996, pp.100-108.

26. Hockey, S. 'Representing humanities text.' Available at: *http:www2.nas.edu/CSTBWEB* (visited October 1998).

27. Mohnkern, K. 'Visual interaction design: beyond the interface metaphor.' Available at: *http://www.acm.org/sigchi/bulletin/1997.2/vid.html* (visited March).

28. Myers, B.A. 'A brief history of human-computer interaction technology.' *Interactions*, March/April 1998, pp.44-54.

29. Shackel, B. 'Human-computer interaction – whence and whither?' *Journal of American Society for Information Science* 48(11), 1997, pp.970-986.

30. Marchak, F. 'Universal design.' *SIGCHI* 30(2), April 1998. Available at: *http://www.acm.org/sigchi/bulletin/1998.2/vid.html* (visited March 1999).

31. Lynch, P.J. 'Visual design for the user interface: part 1 – design fundamentals.' *Journal of Biocommunications* 21(1), 1994, pp.22-30.

32. Vonder Haar, K. 'Visual interaction design: the orchestration age.' Available at: *http://www.acm.org/sigchi/bulletin/1998.1/vid.html* (visited March 1999).

33. Ahlberg, C. and Shneiderman, B. 'AlphaSlider: a compact and rapid selector.' *Proceedings ACM Computer-Human Interaction*, 1994, pp.365-371.

34. Ahlberg, C. *et al.* 'Dynamic queries for information exploration: an implementation and evaluation.' *Proceedings ACM CHI'92*, 1992, pp.619-626.

35. Spence, R. *Information visualisation – a tutorial.* Unpublished notes presented at ICIS Technology Showcase, Nanyang Technological University, Singapore, 14 August 1998.

36. Liew, C.L. and Foo, S. *A study of the application of information visualisation techniques to enhance user's interaction with electronic documents. Technical Report #03/99.* (Singapore: School of Applied Science, Nanyang Technological University, March 1999.)

37. Reed, M. and Heller, D. 'Olive: online library of information visualization environments.' Available at: *http://www.otal.umd.edu/Olive* (visited January 1999).

38. Shneiderman, B. 'The eyes have it: a task by data type taxonomy of information visualizations.' *Proceedings IEEE Symposium on Visual Languages*, 1996, pp.336-343.

39. Card, S.K. 'Visualising retrieved information: a survey.' *IEEE Computer Graphics and Applications* 16(2), 1996, pp.63-67.

40. Lim, E.P., Chowdhury, G,G., Ng, W.K., Hamwamdeh, S., Goh, A. and Foo, S. 'Digital Library for Creative Learning and Research.' Project proposal submitted to National Science and Technology Board (Singapore) by School of Applied Science, Nanyang Technological University, December 1998.

## Notes

i. Bederson, B.B. and Hollan, J.D. 'Pad++: a zooming graphical user interface for exploring alternative interface physics.' *Proceedings User Interfaces and Tecnology*, 1994, pp.17-27.

ii. Rennison, E. 'Galaxy of news: an approach to visualizing and understanding expansive news landscape.' *UIST*, 1994, pp. 3-12.

iii. Mackinlay, J.D. *et al.* 'An organic user interface for searching citation links.' *Proceedings ACM CHI*, 1995, pp.67-73.

iv. Nation, D. *et al.* 'Visualizing web sites using a hierarchical table of contents browser: WebTOC.' Available at: *http://www.cs.umd.edu/projects/hcil/Research/1997/webtoc.html* (visited January 1999).

v. Kandogan, E. and Shneiderman, B. 'Using elastic windows for World Wide Web browsing.' *Proceedings ACM CHI*, 1998, pp.189-190.

vi. Mackinlay, J.D. *et al.* 'The perspective wall: detail and context smoothly integrated.' *Proceedings ACM CHI*, 1991, pp.173-179.

vii. Robertson, G.G. and Mackinlay, J.D. 'The document lens.' *Proceedings ACM User Interface Software and Technology*, 1993, pp.101-108.

viii. Robertson, G.G. *et al.* 'Cone Trees: animated 3D visualizations of hierarchical information.' *Proceedings ACM SIGCHI Conference on Human Factors in Computer Systems*, 1991, pp.189-194.

ix. Furnas, G.W. 'Generalized fisheye views.' *Proceedings ACM CHI*, 1986, pp.16-23.

x. Lamping, J. and Rao, R. 'Visualizing large trees using the hyperbolic browser.' *Proceedings ACM CHI*, 1996, pp.388-389.

xi. Card, S.K. *et al.* 'The WebBook and the Web Forager: an information workspace for the World Wide Web.' *Proceedings ACM CHI*, 1996, pp.111-117.

xii. Bier, E.A. *et al.* 'Toolglass and Magic Lenses: the see-through interface.' *Proceedings SIGGRAPH*, 1993, pp.73-80.

xiii. Stone, M.C. *et al.* 'The movable filter as a user interface tool.' *Proceedings ACM CHI*, 1994, pp.306-312.

xiv. Rao, R. and Card, S.K. 'The table lens: merging graphical and symbolic representations in an interactive focus and context visualization for tabular information.' *Conference on Human Factors in Computer Systems*, 1994, pp.318-322.

xv. Tweedie, L.A. *et al.* 'The influence explorer: a tool for design.' *Proceedings ACM CHI*, 1996, pp.390-391.

xvi. Masui, T. 'LensBar – visualization for browsing and filtering large lists of data.' *Proceedings IEEE Information Visualization*, 1998, pp.113-120.

xvii. Hearst, M.A. 'TileBars: visualization of term distribution information in full-text information access.' *Proceedings ACM SIGCHI Conference on Human Factors in Computer Systems*, 1995, pp.55-66.

xviii. Eick, S.G. *et al.* 'SeeSoft – a tool for visualizing software.' *IEEE Transactions on Software Engineering* 18, 1992, pp.957-968.

xix. Koike, H. and Yoshihara, H. 'Fractal approaches for visualizing huge hierarchies.' *IEEE/CS Symposium on Visual Languages*, 1993, pp.53-60.

xx. Roth, S.F. *et al.* 'Visage: a user interface environment for exploring information.' *Proceedings IEEE Information Visualization*, 1996, pp.3-12.

xxi. Heath, L.S. *et al.* 'Envision: a user-centered database of computer science literature.' *Communications of the ACM*, 38(4), 1995, pp.52-53.

xxii. Plaisant, C. *et al.* 'LifeLines: visualizing personal histories.' *Proceedings ACM CHI*, 1996, pp.221-227.

# The introduction of intranets into the newspaper industry[i]

*Jackie Drennan*
*Research and Information Unit, The Guardian/Observer, 119 Farringdon Rd, London EC1*
*Email: Jackie.Drennan@guardian.co.uk*

**Abstract** *An evaluation of the current state of intranet usage in the UK newspaper industry. Librarians and information managers from all the UK national newspapers were contacted to establish which newspapers had, or were intending to introduce, an intranet. Those who had an intranet were interviewed. To provide a point of comparison with newspapers in the United States, who appeared to have quickly embraced intranet technology, email questionnaires were distributed via the Special Library Association mailing list. The paper shows that librarians are closely involved in the planning of newspaper intranets and describes the management issues, attractions and problems associated with them.*

## Introduction

The advent of Internet technology heralded the possibility of making information available to as wide an audience as possible; newspapers have exploited these new opportunities to branch into different markets and reach new readers by creating web versions of, or parallel publications to, their print editions. *The Electronic Telegraph*[1] and *Guardian Unlimited*[2] are examples of the diversification of newspapers into this new medium.

Since 1995, organisations have exploited intranet technology in a different manner – to create intranets which operate as secure, internal information and communication systems. The take-up of intranet technology has been rapid – Bonnie reports that of the top 200 UK companies, 48% have an intranet, and 77% of those that do not are planning to introduce one[3].

The newspaper industry would appear to stand to benefit greatly from intranet technology. Journalists' demands for large amounts of up-to-date, accurate, information, and the speed with which they need access to this information, seem to make the newsroom an ideal candidate for an intranet. A secure internal network could transform the way in which information is stored, manipulated, distributed and retrieved, meeting the needs not only of the journalistic staff but all employees.

Newspapers in the United States have been experimenting with introducing intranets. At a conference of the SLA (Special Libraries Association) News Division[4], information professionals who had worked with intranets in their newspapers gave these reasons for building a newsroom intranet:

- reaching bureau reporters and others outside the newsroom editing system;
- navigational aid to the greater Internet;
- gateway to in-house resources.

US news librarians have been keen to share their experiences of intranets; several of them have placed examples of their intranet onto the Internet, to assist fellow librarians considering introducing one[5]. The intranets range from very simple front-ends, to more sophisticated graphics and Java programs, but behind these different styles, the purpose of the intranet is the same – to assist journalists' news-gathering by providing timely, relevant, accurate information in an easy-to-use format.

In the UK, newspapers are starting to take note of the possibilities of newsroom intranets. Several already have functioning intranets, while others are in the process of creating one. This study examined the current state of intranet usage in the UK newspaper industry. The librarians or information managers of all the national newspapers were contacted to establish which newspapers had, or were intending to introduce, an intranet; those who responded that their paper did have an intranet were interviewed. To provide some comparison with newspapers in the United States, who appeared to have embraced intranet technology more rapidly, email questionnaires were distributed via the SLA mailing list; five newspapers replied.

### Prevalence of intranets in UK newspapers

Of the nine UK newspaper organisations contacted, five have an intranet, one is developing an intranet, another plans to introduce one in the future, and two have no plans for an intranet (figures correct August 1998).

Within those newspapers which have intranets, there is wide variation in the scope of the systems and the uses to which they are put. The intranets can be grouped into three main categories:

1. a vehicle for delivering the newspaper's electronic text archive to end-users in a web-browser format (two newspapers);
2. an editorial intranet, set up and maintained by the library or information department, containing selected web-links, pre-prepared fact-files and chronologies, and access to the electronic archive (three newspapers);
3. a company-wide intranet, with sites for all departments including editorial, research, personnel and production (one newspaper).

At present, it appears that the intranets being introduced into newspapers are specialised in their function: the main or original purpose of all of them is to provide information for journalists to assist with the writing of articles and features, and the majority of the intranets do not cater for the whole staff of the company or contain company information. This situation looks likely to change in the future, as several newspapers are contemplating creating a company-wide intranet.

### Development of newspaper intranets

All of the UK newspapers surveyed said that the electronic production of the newspaper had pre-empted or facilitated the creation of an intranet, as the data from the paper could be captured and stored to create a database. In the case of four of the newspapers, the suggestion to create an intranet had originated in the library or information unit (in the other two newspapers it was the IT department who initiated the intranet) – one of these was directly inspired by the US sites on the SLA website.

Surprisingly perhaps, none of the UK newspapers had carried out an information needs assessment of journalists before establishing the intranet, though two of the US papers had done so via questionnaires. All the UK libraries felt that they already knew the needs of their customers[ii] and some expressed the opinion that it would be futile to ask users what they would wish to see on the intranet, as it would be too great a task, or that journalists would not know what they wanted. While it may be true that newspaper libraries have been catering to the needs of journalists for so long that they can anticipate the kind of questions that will be asked, and recognise patterns of research, the introduction of an intranet offers possibilities for new and dynamic means of information provision. Nicholas has noted that information professionals have largely neglected undertaking needs assessments in the past[6] and there is a danger that, by basing the information offered only on what has always been provided, libraries may not be exploiting the opportunity to develop new services in response to new needs.

### Management of newspaper intranets

The issue of intranet management has not yet become a major issue in the UK newspapers surveyed, because in most cases the intranet is still in its infancy, and consists of information created and maintained by one department, either the library or IT services. The literature on corporate intranets suggests that as systems grow, the issue of management will need to be addressed to prevent chaos ensuing, and to ensure that information remains accurate and accessible[7].

Gupta and Wachter[8] suggest four methods of intranet management:

1. Laissez-faire, or decentralised model. All users are free to put on whatever information they choose, with a minimum of bureaucracy.
2. Centralised model. Policies and procedures are implemented to control development.
3. Mixed model. Policies and procedures are handed down throughout the organisa-

tion, but it is up to the individual user to use intranet technology within these restrictions.

4. Provide support services. Users are supported in their developmental efforts through the provision of services, such as programming in JAVA or CGI.

None of the newspapers surveyed has adopted, or intends to adopt, an entirely centralised model of management, whereby any department wishing to publish information must submit a request to the appropriate intranet staff. In cases where the library has control over the intranet, staff reject this idea, fearing that they could become a 'dumping ground' for other departments' work. In cases where the IT department has overall responsibility, they do not have responsibility for creating pages for departments, only checking technical details, and control is still decentralised. Newspapers are taking advantage of the nature of intranets as a distributed information system, by not restricting information provision to one department only. Gupta and Wachter's 'mixed model' seems to be the solution which UK newspapers are either adopting or plan to adopt once the intranet becomes company-wide.

Guidelines as to content and design will be handed down through the company, but departments will be free to contribute information within these boundaries. Other management issues, beyond determining how the intranet develops, need to be considered. Even if the intranet remains within the bounds of the newsroom, establishing policies for responsibility for, and maintenance of, information are important. Because the information distributed via a newspaper intranet (or the editorial section of the intranet) is intended for use and reproduction by journalists and subsequent publication, accuracy is paramount. Publishing inaccurate information in the newspaper can lead, at best, to the lowering of the paper in readers' esteem and at worst to legal proceedings if the inaccuracies constitute libel. Only one of the newspapers surveyed, however, saw the issue of 'provenance' or liability for inaccurate information to be a serious concern. While it is true that newspaper libraries are used to meeting journalists' demands for accurate and timely information, distributing information via an intranet creates new responsibilities. The librarian is no longer acting as a gatekeeper to information, checking and verifying sources as and when information is provided in response to a request, but is rather providing users with a constant source of potentially useful information. This may be drawn upon at any time by staff in the faith that it will be current and accurate and the burden is on the workers responsible for the intranet to ensure that it is so.

At present in the UK newspapers, this responsibility is devolved between the members of staff who place the information on the intranet, and no individual has overall control. The situation among US newspapers would appear to be slightly different – only one paper said that 'the staff placing the information on the intranet' have responsibility for ensuring its accuracy – in two of the papers information staff are responsible, in another the intranet manager (an assistant managing editor) takes responsibility, and in another case all material has to be approved by the library or systems department. These increased levels of double-checking could be due to the higher prevalence of litigation in the US than in the UK, and thus higher awareness of possible liability for inaccuracies, or simply be due to the fact that the intranets have been established for longer, and more of them are company-wide.

None of the UK newspapers has a position of 'intranet manager', though at two of the newspapers one member of the library staff has general responsibility for the intranet. In the papers where the IT department oversee the intranet, there is still no individual 'intranet manager'. Among the US papers, two of the five have intranet managers, but in another two cases the library staff are the nearest thing to an intranet manager. The possibilities for information staff to assume positions of management with regard to intranets are clear; while no strategy for management exists, it is information professionals, who are used to managing information as effectively as possible, who can provide direction. Decisions need to be made about intranets, and information professionals are well placed to make them if they seize the opportunity.

### Benefits of newspaper intranets

None of the newspapers interviewed, whether UK or US, cited financial benefits as an advantage of their intranet, although Zorn found that this was one of the reasons for their popularity[9]. Though one UK paper is looking for ways to make their intranet profitable, through incorporating financial and market information, and possibly fostering new business links through an extranet, none of the other participants in the study mentioned financial implications of the intranet. The fact that all of the interviewees and respondents were employed in library and information sectors, rather than the financial or managerial side of their newspaper may well have had some bearing on these results, but the intranets which have been developed in newspapers appear to be designed to deliver benefits in terms of communications and information flow rather than cost.

The fact that intranets are platform independent is a benefit in a newspaper environment. In at least two of the UK newspapers interviewed, staff use both Macs and PCs, so any information system must be able to be supported on both systems.

The ability to link in 'legacy' systems and databases was not such an issue as might have been expected. In some cases, the arrival of the intranet coincided with the introduction of a new computer system and the intranet is not available through the previous editorial system. Though all of the newspapers use the intranet to provide an interface to their text database, and in the case of the UK papers, to the material of the data exchange, the intranet is not used to provide access to other, commercial, databases such as FT-Profile. The use of these databases remains the province of the library, because of the costs involved.

Rather than benefits in terms of hardware or costs, newspaper intranets seem to offer benefits in terms of improved access to information. Presenting journalists with a wider variety of information sources, in an easy-to-access format, from their own desk seems to be the main drive behind the editorial intranets studied. Empowering users to access information for themselves not only saves their time, but also that of the library, who benefit from reduced numbers of low-level enquiries such as spelling checks. The four UK newspapers who create files of reference information for this specific reason all hope to be, or have already been, able to devote more time to more in-depth research, more appropriate to their professional status than checking spellings or house-styles. Creating a more information-rich environment is the greatest benefit of newspaper intranets at present.

### Problems with newspaper intranets

Apart from specific technical complications, the time required for keeping material on the intranet current and accurate is the most common problem faced by newspapers. When material is compiled and placed on the intranet, it needs to be maintained, and the very nature of the information used by newspapers – current affairs – means that it can quickly become out-of-date or inaccurate.

The prospect of expending such a lot of staff time on updating, and the possibility of information becoming out-of-date has been enough to deter two newspapers from placing files of information onto the intranet. Other newspapers are trying to overcome the problem by establishing schedules for updating, allocating specific staff to the task, or using web-links to replace files of information wherever possible. The fact that the US newspapers still find time demands problematic suggests that finding a solution will not be easy, and the problem may in fact grow, as the amount of material on the intranet grows.

Though none of the newspapers specifically mentioned information overload, one US paper did comment that the intranet contained 'so many resources and great information that the reporters and editors don't know what they have', and that 'they get lost and don't take advantage of everything'. Some UK papers have also said that there are limits to the intranet – as one librarian commented, 'you can't start rebuilding all the reference books in the library'. To strike a balance between providing sufficient information and keeping material at a manageable level will be a challenge faced by intranet managers. It is an area in which user needs analysis could help in decision-making; users know which types of information they find most useful, and intranet developers can build on this knowledge to create systems of the most relevant material. Creating high quality material and services that are valued and useful to staff, rather than placing as much material as possible onto the intranet and not being able to maintain it, will avoid the type of situation envisaged by one UK librarian where the intranet was thought to be so shallow as to be of little use to anyone.The fact that none of the intranets studied has a search engine for the whole intranet has not yet caused problems, as the intranets are small in scope. Some newspapers have increased searchability by placing information in databases, each with their own search engines. Though this suffices at present, once more and more material is stored on the intranet, users will find it frustrating not to be able to search the whole of the intranet for a specific piece of information.

### Newspaper librarians and intranets

It was encouraging to find that newspaper libraries were involved in all of the UK intranets. All of the libraries had been involved in the planning stages, which as Adourian and Schweyer[10] point out, is vital if the library wishes to be a part of the project.

In four of the UK papers, and all of the US papers, librarians are acting as web authors, creating pages for the intranet and all are, to some extent, content analysts as they are making decisions about what would be most useful for journalists on the intranet. Though some are already acting as trainers for the intranet, others have not yet adopted this role, and this seems to be an ideal area for librarians to move into. Many have already trained journalists to use the Internet, or the newspaper's text database, so training for the intranet seems to be a logical extension of this. Another role which librarians have not yet taken on is that of editor and evaluator, critiquing other departments' web-sites. Several libraries have plans to do this and others already do so in an unofficial capacity.

The big challenge for news librarians is the move into the area of web administrators, overseeing the control of the intranet. At three of the UK newspapers, the library is already playing this role, as they are the only department placing material on the intranet, but it remains to be seen whether they will continue to hold this role once the intranet becomes company-wide. The fact that at one US paper, the intranet was started by the library, who then developed it into a corporate-wide intranet and continue to be in charge of the whole project shows that it is possible for libraries to hold this position of responsibility.

The US newspaper library which manages the entire corporate intranet found that this role made them more visible outside the newsroom, and made them appear more 'technically savvy'. Not all of the UK libraries were convinced that being involved with the intranet would raise the status of their service, but this is not as negative as it might appear. Those who did not think that their status would be raised already felt that the library enjoyed good standing within the company, and pointed out that an intranet alone would not improve their profile, but that the whole range of services provided were important. Several did, however, feel that if users had a system on their desk that had been designed by the library, this could not help but raise the department's profile. This is why 'badging' the intranet pages with the name of the library is popular, as it is a form of marketing for the library. The area in which the libraries felt their status would be most improved was technology; that librarians were not just involved in cutting papers, but were also capable of exploiting new technologies. This is important if newspaper libraries are to escape their image as being solely the paper's archive, or 'morgue' and be seen as more proactive and professional. Moving into the corporate side of the intranet, and dealing with information that has, typically, not been the library's domain – such as market reports and financial information – will do more to raise awareness of the library and transform its profile. This would, by necessity, involve working with departments such as Marketing, who do not normally use the library service and thus raise awareness of the potential of the library and possibly provide a new customer base.

## References

1. *The Electronic Telegraph. http://www.telegraph.co.uk/*

2. *Guardian Unlimited. http://www.guardian.co.uk*

3. Bonnie, L.N. *The use and non-use of intranets in commercial organisations.* MSc dissertation, Department of Information Science, City University, 1997.

4. Basofin, P., Jesse, M. and Krueger, S. 'Building a newsroom intranet: gaining visibility through mastering technology. SLA conference, Seattle, 8 June 1997. *http://sunsite.unc.edu/slanews/conferences/1997/CE/*

5. *http://sunsite.unc.edu/slanews/intranets*

6. Nicholas, D. Assessing information needs: tools and techniques. (London: Aslib,1996.)

7. Wiener, S. and Hanson, K. 'Policies and guidelines: rules of the road for an intranet.' *Intranet Journal. http://www.intranetjournal.com/expert/rules.htm*

8. Gupta, J. and Wachter, R. 'The establishment and management of corporate intranets.' *International Journal of Information Management* 17(6), 1997, pp.393-404.

9. Zorn, P. *et al.* 'Surfing corporate intranets; search tools that control the undertow.' *Online* 21(3), May 1997, pp.31-51.

10. Adourian, A. and Schweyer, K. 'Intranets and the proactive librarian.' *Information Outlook* 1(17), 1997, pp.19-22.

## Notes

i. Based on an MSc Information Science dissertation (City University). Awarded the FTProfile prize for the best dissertation produced in 1998.

ii. *The Guardian* in fact had done a number of user needs analyses, although not directly related to the intranet.

# Building the digital library on solid research foundations

*Ian Rowlands and David Bawden*
*Department of Information Science, City University, Northampton Square, London EC1V 0HB*
*Email: ir@soi.city.ac.uk*

**Abstract** *The digital library is a socio-technical concept of great significance. It redefines the relationships between information providers and intermediaries and, potentially, transforms the way that services are delivered to users. This article, based on a British Library Research & Innovation Centre funded study, reviews current themes and directions in digital library research and scholarship. It locates the digital library in a simple work-oriented framework emphasising its social as well as its systems and informational dimensions. The article highlights differences in understanding of the digital library construct between the library and computer science communities and identifies some critical areas for further research.*

## The scope of digital library research

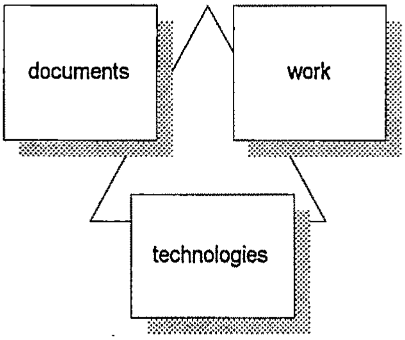
The term 'digital library' is a source of much debate and confusion[1]. In this article, we take the term to mean a distributed library information service, located either in a physical or a virtual space, or a combination of both, in which a significant proportion of the resources available to users of that service exist only in digital form. We thus distinguish the digital library from the 'electronic library', which simply provides access to a range of material in digitised form within the framework of a traditional library, and 'hybrid' or 'gateway' libraries, which lie somewhere on a continuum between the traditional and the digital library, with electronic and paper-based sources integrated and used alongside one another. Proponents of the hybrid library regard it as a worthwhile model in its own right, others merely as a transitional state between traditional and digital libraries.

Research into the digital library is scattered across a wide range of disciplines and published sources. Our initial review of the literature revealed eight thematic clusters:

- human factors;
- organisational factors;
- library management;
- futures studies and scenarios;
- information law and policy;
- systems factors;
- knowledge organisation and discovery;
- impacts on the information transfer chain.

These clusters reflect contrasting approaches to the study of digital libraries and the multidisciplinary nature of the research questions posed. The conceptual framework, which we develop here, recognises that diversity but adopts a work-oriented perspective based on the work of Yates[2].

*Figure 1: Yates' work-oriented library model.*



Yates highlights three crucial aspects, which may be applied to our understanding of both traditional and digital libraries. Libraries, however much they differ outwardly, have one thing in common: they house and provide access to stores of *documents*, defined in the broadest sense to include enduring communicative records. These documents are created and maintained using *technologies* that have required considerable human investment to develop – from the printed book to the object-oriented database. And both documents and technologies are deployed through the institution of the library to support *work* – both the work of the researchers who use the library and of the library's own staff who provide reader and other services.

This simple framework emphasises the interconnectedness of documents, technologies and work, and stresses that innovation in the digital library cannot be soundly based on advances in a particular technology alone. Other writers seem to have reached a broadly similar conclusion:

> 'Digital libraries are best thought of as human resource streams that unite readers, authors, librarians and researchers with electronic materials, resource streams, computer equipment and know-how.'[3]
>
> 'By stressing the importance of either technology or librarianship, commentators on the digital library have created a false dichotomy; digital libraries need librarians in partnership with domain and information technology experts.'[4]

In a 1995 paper, Daniel Levy and Catherine Marshal[5] think through some of the consequences that flow logically from Yates' model when applied to the digital library. They point out, for example in relation to the Internet, that technical infrastructure does not, in and of itself, constitute a digital library. That requires content and, more particularly, managed collections of content for identified user communities.

For the purposes of this report, the three elements of Yates' model have been re-labelled as the *informational* (documents), *systems* (technology) and *social* (work) domains of the digital library. These domains offer a more meaningful grouping of the thematic clusters identified earlier, which are now more clearly seen to be overlapping.

***Social domain***
- human factors (non-machine specific)
- organisational factors
- library management factors
- information law and policy factors
- impacts on the information transfer chain
- futures studies and scenarios

***Informational domain***
- knowledge organisation and discovery (e.g. metadata)
- impacts on the information transfer chain (supply-side activities)
- futures studies and scenarios

***Systems domain***
- human factors (e.g. human-system interaction)
- knowledge organisation and discovery (e.g. software agents)
- impacts on the information transfer chain (technical)
- futures studies and scenarios

In the sections that follow, an attempt is made to summarise the key research issues evident in recent scholarly and professional literature on the digital library. Paradoxically, given the pre-eminent place accorded to information technologies in the literature, we begin with social factors, in the belief that many current digital library developments unduly emphasise questions of what? and how? rather than why? Perhaps we need to focus more directly on the collaborative nature of research and service delivery in the digital library and thus regain a fuller sense of the workplace activities and processes that digital libraries should be supporting.

## Research in the social domain

As might be anticipated, research into the social contexts surrounding the digital library (here defined broadly to include psychological, cultural, organisational, economic, legal and policy elements) is characterised by high disciplinary and literary scatter. Research in the social domain is considered here under three main headings:
- information skills and literacy;
- impacts on work and organisations;
- information law and policy concerns.

### *Information skills and literacy*

The advent of the digital library carries with it a need to re-assess what the notion of 'literacy' really means at this point in history. The term carries a variety of meanings and considerable overlap with other terms, such a 'library literacy', 'network literacy' and, most recently, 'digital literacy'[6]. It is best understood, however, as the complex of skills and attitudes involved in accessing, evaluating and using information, especially where that information is obtained from a variety of sources and in a variety of formats. Such skills are clearly needed in the complex environment of the digital library. Conversely the digital library environment is a rich arena for developing such skills.

A consensus seems to be emerging that learning in digital library environments encompasses two distinct components: learning how to access, evaluate and use information resources; and learning now to master and build on the ideas embodied within those resources. In its most highly developed form, information literacy has been characterised by 'free form learning and just-in-time information delivery'[7].

The phrase 'information literacy' is peppered liberally throughout the library and information science literature, although the concept is not at well articulated. This is not to say that the concept has not been worked on – there is a surprisingly large body of empirical and conceptual work in the educational literature which relates information technologies and, increasingly, networked information resources to learning outcomes. As national policy frameworks for the digital library reach out to embrace the citizen and the concept of lifelong learning, there is a pressing need for this rapidly accumulating body of knowledge to be incorporated into the design and especially the evaluation of digital library pilot and demonstration projects.

### *Impacts on work and organisations*

An assumption, sometimes explicit, more usually implicit, informing the design of the digital library is an idealised form of use – the lone researcher sitting at a workstation. The origins of this (mis?) conception probably derive by extension from client-server architectures and the ordered world of systems thinking. Yet we know from observation and personal experience that information seeking, research activities and library services are often highly collaborative. Ehrlich and Cash, for example, note that

'People seek information by communicating with members of their communities; they look not only for materials and specific answers, but for corroboration, new interpretations, and new methods of finding information... all this suggests that support for communication and collaboration is just as important as support for information seeking activities and that, indeed, support for the former is needed to support the latter.'[8]

The workplace implications of the digital library are profound. Digital technologies relax the constraints of time and place in relation to how people work with one another. New organisational forms are beginning to emerge, such as collaboratories, but the ultimate effects of these new ways of working are unknown. While some advocates of the digital library see the elimination of the library as 'place', this now seems an increasingly extreme – and technologically determinist-position to adopt. More research is needed that focuses on user communities, possibly using ethnographic techniques, to observe the impact of electronic library infrastructures on working practices.

The impact of the digital library on existing institutions and institutional structures is another focus of research interest, not least in relation to the future role and place of traditional libraries. Positions on this issue range from those held by 'transformists' who foresee the disappearance of the physical library and its phoenix-like remanifestation in a completely virtual form, to 'continuists' who stress resistance to change, inertia and 'more of the same'. The critical factor in this equation is people:

> 'Digital library proponents must consider the role of people (as users and service providers) if the digital library is to be truly beneficial. Technology and information resources, on their own, cannot make up an effective digital library.'[9]

Methods for selection, acquisition, organisation, access and preservation in the digital library differ considerably from those in traditional libraries and models of best practice urgently need to be developed and promulgated. The focus among library consortia is gradually shifting from integrated library systems to digital library projects, resource sharing, collective database licensing and document delivery, each of which implies considerable changes to working practice. Indeed, some of the more alarmist pronouncements on the digital library foresee the elimination of intermediaries (searchers, librarians, retailers, distributors and others) who merely 'interfere' in the process of interchange between content creators and readers.

The culture change implied goes beyond different routines or particular tasks; rather it implies a far-reaching change of ethos and values, reflected in a need for new skills, knowledge and attitudes. The general presumption seems to be that users will become more self-reliant in their information seeking and using behaviour, while librarians will become providers of systems and interfaces rather than resources *per se*. It may be, however, that this view reflects the fact that most digital library research has been carried out within the context of higher education.

Finally, the fragility of cultural memory in a digital age should not be underestimated. This aspect of the digital library is a source of concern for many commentators:

> 'The difficulty and expense of preserving digital information is a potential impediment to digital library development. Digital preservation is largely experimental and replete with the risks associated with untried methods. Digital preservation strategies are shaped by the needs and constraints of repositories rather than users, current or future.'[10]

The term 'digital preservation' seems to have two somewhat distinct meanings: the preservation of materials created in digital form; and the digitisation of analogue materials for access and preservation. We take digital preservation here to mean the processes by which a digital library ensures the long-term viability of its digital materials, whatever their source, by protecting them from the threats of physical deterioration and technological obsolescence. It is a problem that many libraries with some digital components are only now beginning to encounter[11].

### *Information law and policy concerns*

The migration of information resources to an increasingly digital library infrastructure throws up many legal and policy issues. These relate, *inter alia,* to information integrity, document authenticity, user confidentiality, payment systems, security encryption, personal privacy and, not least, to copyright management.

Confusion over intellectual property rights in the digital age, the profit motive, the concept of fair use and the development of the global economy have converged to create a vexingly complex set of problems. Given the variety of stakeholders involved, it is clear that we will have to extend the principles of copyright and develop new concepts of ownership and archiving. Or at least that is the prevailing wisdom. Some commentators argue that the challenge is so great that no amount of tinkering can remedy the problem. Instead we need a fresh start, one designed from the ground up:

> 'Copyright problems cannot possibly be solved within nonlocalised cyberspace. Real libraries, with their inherent local limitations, constitute the only means our society has for overcoming the 'what' and 'who' restrictions of virtual libraries.'[12]

Authentication is a closely related issue. By authentication, we mean the processes and procedures by which user access to information resources is controlled in the digital library environment. Such measures are needed to protect the commercial interests of rights holders and the personal privacy of the users themselves. Authentication is more precisely the establishment of the identity of the user in a particular context, while authorisation controls levels of permitted access to resources for that user identity[13]. In a traditional library this is a relatively straightforward process, usually involving some form of 'library card' and the control of physical access to a building. In digital library environments the potential for unauthorised access is clearly much greater, but the widely noted phenomenon of 'password proliferation' has itself effectively become a barrier to user access.

In a very real sense, digital libraries are evolving in a legal and policy vacuum and against a backdrop of commercial and market uncertainty. Libraries embody many attributes, not just the satisfaction of narrow information seeking behaviour. They have an important social function, providing a focal point in organisational, local, and even national life. This aspect is highly significant for national library policy – citizen research from Japan and the USA, for example, suggests that there is as yet no public consensus around virtualisation of their libraries. Indeed there seems to be a growing feeling that electronic libraries are being developed contrary to the needs and wishes of the public[14]. This disquiet is amplified by Collins:

> 'The library world is in the throes of a drama involving conflicts and contradictions between Gutenberg pasts and McLuhanesque futures, centering on the demise of the library as a social institution. The digital library presents information in a way which is remarkably free of conflicting interpretation and historical construction; it just 'is'. By suppressing the social dimension of information transfer, we threaten to delegitimise the library as a public institution.'[15]

More policy-focused research is needed to establish the public's real, as opposed to their presumed, requirements from digital library infrastructure; to understand how digital library initiatives fit in with broader social and economic policy objectives; and, at the institutional level, to identify more effective strategies for managing change.

While there is a need for focused research into specific issues such as rights management, data privacy and payment systems, the bigger picture, what Sandra Braman calls the 'policy regime' also needs to be addressed[16]. A policy regime is an overarching set of principles and values, which provides a common framework of understanding. Only within such a regime may a fully coordinated and effective set of policies be developed. Policy research into the normative structure of information policy and its optimal framework for the digital library is badly needed. While we cannot put aside immediate practical problems until these larger theoretical issues have been resolved, the recent work of the Library & Information Commission in developing elements of a national information policy[17] should be encouraged and more widely disseminated within the digital library community.

**Research in the informational domain**

Information retrieval in distributed network environments poses many challenges. The most notable is the degree to which it can be made to operate effectively at the level of meaning (semantics) rather than character matching (syntactics) across large heterogeneous document collections. 'Deep semantic interoperability' is a phrase that often crops up in digital library vision statements, but precisely how we get there is a far from trivial question, especially given the sheer scale of many digital library applications. Human indexers cannot possibly hope to keep up with the influx of digital information, nor can they be sufficiently consistent. We need uniform and automatic mechanisms – 'extractors' – to build meaningful and expressive indexes with which expert novice users can interact:

> 'Information exploration in digital libraries involves building bridges between document space (the realm of information artifacts, ranging from books to mails to bibliographic references) and concept space (the context and knowledge required to interpret documents).'[18]

'Information exploration' is a complex notion embracing a number of more familiar concepts: accessing, searching, browsing and navigating for information[19]. Searching generally implies purposive behaviour and the identification of relevant items by subject or other criteria that are clear to the user. Browsing and navigating usually, but not invariably, imply a more loosely structured approach. Although more detailed typologies of searching and browsing are described in the information retrieval literature they do not seem to have been studied systematically in the context of the digital library.

This is surprising when one thinks of the nature of the digital library – a network of distributed repositories where information objects of any type can be searched and retrieved using technologies that can transparently handle variations in protocols and formats. If this vision is to become a reality, information objects in the digital library need to be fully documented with contextual information relating to their source creation, processing and current status, as well as by flags signifying their content. The form this metadata (information about information) should take, and the standards that are ultimately the most widely accepted, has become a critical issue in the evolution of the digital library.

The term metadata was, apparently, first used in the literature in 1962, and made its premier appearance in a dictionary in 1980[20]. Metadata is a contentious issue because its function is perceived very differently by librarians and computer scientists. The research literature reveals a polarisation between the librarian's bibliographic control approach to metadata and the computer scientist's data management approach:

> 'MARC cataloguing and the OPAC are in collision with the world of the Internet. Librarians are to blame; they initially isolated themselves from developments in the world wide web, especially the notion of universal access, and they de-emphasised 'ephemeral' resources. It isn't too late for a merger of the best of both worlds – the full text analysis provided by search engines and the controlled vocabularies found in library OPACs.'[21]

Traditional cataloguing practice differs fundamentally from the rich metadata needed to describe the materials in a distributed digital library and three functional categories of metadata have been proposed: metadata at the descriptive (or intellectual), structural and administrative levels[22]. The MARC format beloved of the library community really only deals well with intellectual metadata and so a more inclusive concept is needed to merge these various perspectives. Metadata standards are beginning to emerge that may become to digital libraries what MARC has been, and still is, to print libraries, the most notable current contender being Dublin Core[23].

Evaluating the performance of information retrieval systems in highly distributed environments is another area where research is needed. It is becoming increasingly clear that precision and recall are far from adequate measures of retrieval effectiveness in hypermedia digital library environments. New metrics, perhaps based on the structural analysis of hypermedia networks and the browsing and navigation behaviour of users need to be developed and tested[24].

## Research in the systems domain

Two big research and development issues stand out in the systems domain: interoperability and scaleability. Coordinated progress in both areas is needed to deliver Hunt and Ethington's vision of digital libraries that can offer:

> '... full realization of the potential for infinitely integrated growth in materials, genres and formats.'[25]

The issue of interoperability permeates almost all aspects of the digital library. Interoperability has implications for systems architectures, standards and protocols for data transfer, metadata and markup languages. Interoperability is a fundamental goal in constructing digital libraries from independently developed collections that rely on each other to accomplish larger tasks. Ideally, these building blocks would be capable of evolving independently yet at the same be able to call on one another efficiently and conveniently[26].

Various mechanisms are used to maximise platform and software independence (e.g. entities, link protocols). Though often effective they may be used incorrectly or in ways which are difficult to implement. Many difficult questions remain for libraries seeking interoperability[27]. Particular concerns have been voiced in the literature about the perceived gap between the institutional needs of libraries and the products offered by library systems vendors and this is an area that warrants further exploration[28]. The concept of interoperability reflects, in outline, the need for information in disparate formats and media to be accessible by users through a common interface with a single authentication process. In its most complete form, this would lead to 'seamless' integration. Its antithesis is the situation observed not infrequently today, in which a user, presented with some information on the screen, scribbles it down onto paper, before keying it into another system. The problem is one of fusing together heterogeneous components: different computer systems, operating environments, storage media, information formats, processing and quality standards, subject descriptions and the like into a coherent service.

The notion of interoperability therefore summarises a variety of issues that require integration at a number of levels, from physical connection, through standards for format interconversion and resource naming, to the semantic aspects of metadata resources. Digital libraries of the future will inevitably be shaped by the technologies, standards and models adopted today. Interoperability is therefore a major issue[29].

The scaleability of digital libraries, potentially to national and international dimensions, is only possible within the limits set by interoperability standards (or the lack of them). Schatz and others argue that size really does matter when it comes to digital library research:

> '...attention to scaling as a research theme is essential and may help in further clarifying infrastructure needs and priorities ... there is a growing consensus on the need to enable large scale deployment projects (large in terms of user community, numbers of information objects, numbers of repositories) and subsequently to fund studies of the effectiveness and use of such systems. It is clear that limited deployment of prototype systems will not suffice if we are to fully understand the research questions involved in digital libraries.'[30]

Multimedia digital libraries involve different types of data objects, many of which are large and may cause transmission delay when accessed in a distributed environment. Optimising the trade-offs between response time and the quality of the decompressed output data is another critical technical issue in user acceptance.

## The future of digital libraries?

There are many competing visions and much uncertainty over what the digital or virtual library will or should look like. One thing that is certain is that the digital libraries of the future will be shaped by the technologies, standards and models adopted today. The question is are we being bold enough? Some commentators believe that the current status of digital library research and development is far too conservative:

> 'Discussions of the digital library usually emphasise the transition of scholarly information and libraries to the electronic medium. To that extent, the discourse is framed in the context of the print world we all understand. Most of the attempts to build 'digital libraries' build on these analogies. On the other hand, it is likely that this thinking about 'transition' is itself transitional and merely looks at the past as a way to conceptualise the future. Such thinking may constrain opportunities in the new networked world.'[31]

In other words, certain features of existing libraries are being unreflectively conserved, as are some features of an idealised and unreal past. The future surely lies in developing understanding between librarians, technologies and users:

> 'To participate most fully, librarians need the help of technologists to better understand the possibilities being created by digital technologies and technologists will need the help of librarians to appreciate the richness of traditional librarianship and to identify aspects of it that are most relevant to the continuing evolution of libraries.'[5]

## References

1. Bawden, D. and Rowlands, I. *Understanding digital libraries: towards a conceptual framework.* Research Report 170. (London: British Library Research & Innovation Centre, March 1999.)

2. Yates, J. *Control through communication.* (John Hopkins University Press, 1989.)

3. Covi, L. and King, R. 'Organisational dimensions of effective digital library use: closed rational and open natural systems models.' *Journal of the American Society for Information Science* 47(9), 1996, pp.672-689.

4. Matson, L.D. and Bonski, D.J. 'Do digital libraries need librarians? An experimental dialog.' *Online* 21(6), 1997, pp.87-92.

5. Levy, D.M. and Marshall, C.C. 'Going digital: a look at the assumptions underlying digital libraries.' *Communications of the ACM* 38(4), 1995, pp.77-84.

6. Bawden, D. *Information literacy and other literacies.* Paper in draft, 1999.

7. Weinberger, M.I. 'Just-in-time learning with electric Library.' *Library Trends* 45(4), 1997, pp.623-638.

8. Ehrlich, K. and Cash, D. 'Turning information into knowledge: information finding as a collaborative activity.' In: *Digital Libraries '94, College Station, Texas: 19-21 June.* pp.119-125.

9. Sloan, B. 'Service perspectives for the digital library: remote reference services.' *Library Trends* 47(1), 1998, pp.117-143.

10. Hedstrom, M. 'Digital preservation: a time bomb for digital libraries.' *Computers and the Humanities* 31(3), 1997, pp.189-202.

11. Hedstrom, M. and Montgomery, S. *Digital preservation needs and requirements. www.rl.org/lpreserv/digpres.html* (visited June 1999).

12. Mann, T. 'Reference service, human nature, copyright and offsite service in a 'digital age"? *Reference & User Services Quarterly* 38(1), 1998, pp.55-61.

13. Lynch, C.A. 'The changing role in a networked information environment.' *Library Hi-Tech* 5(1/2), 1997, p.30.

14. Tsuno, K. *et al.* 'Digital libraries and citizens' rights.' *Toshokan Sushi* 92(5), 1998, pp.336-356.

15. Collins, S.G. *'Knowledge will forever govern ignorance': power and ideas of the Formation Society at the Library of Congress.* PhD thesis. (Washington DC: University of Washington, 1998.)

16. Braman, S. 'Defining information: an approach for policy-makers.' *Telecommunications Policy* 13(3), 1990, pp.233-242.

17. *Towards a national information policy for the UK.* (London: Library & Information Commission, 1998.)

18. Baldonado, M.Q.W. *An interactive structure mediated approach to exploring information in a heterogeneous distributed environment.* PhD thesis. (Colorado: University of Colorado, 1997.)

19. Ellis, D.A. 'Behavioural model for information retrieval systems research.' *Journal of Information Science* 15(4/5), 1989, pp.237-247.

20. Bide, M. *Introduction to metadata.* Paper presented at British Computer Society meeting, London, 10 March 1999.

21. Wilder, D. and Greenfield, R. 'ESS: the Library of Congress experimental search system.' *Library Hi-Tech* 15(3/4), 1997, pp.127-135.

22. Ng, K.B. *et al.* 'Control or management? A comparison of the two approaches for establishing metadata schemes in the digital environment.' *Proceedings of ASIS* 34, 1997, pp.337-346.

23. *Dublin Core and metadata: a tutorial. hosted.ukoln.ac.uk/metadata-1997/tutorial/presentation* (visited June 1999).

24. Salampasis, M. *et al.* 'Evaluation of information seeking performance in hypermedia digital libraries.' *Interacting with Computers* 10(3), 1998, pp.269-284.

25. Hunt, L. and Ethington, P.J. 'The utility of spatial and temporal organization in digital library construction.' *Journal of Academic Librarianship* 23(6), 1997, p.475-83.

26. Paepcke, A. *et al.* 'Predicate rewriting for translating Boolean queries in a heterogeneous information system.' *ACM Transactions on Information Systems* 17(1), 1999, pp.1-39.

27. Cole, T.W. and Kazmer, M.M. 'SGML as a component of the digital library.' *Library Hi-Tech* 13(4), 1995, pp.75-90.

28. Healy, L.W. *Library systems: current developments and future directions.* (Washington DC: Council on Library and Information Resources, 1998.)

29. Tennant, R. 'Digital libraries: the grand challenges.' *Library Journal* 122(20), 1997, pp.31-32.

30. Schatz, B. *et al.* 'A parallel computing approach to engineering concept spaces for semantic retrieval: the Illinois Digital Library Information Project.' *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18(8), 1996, pp.771-782.

31. Lyman, P. 'Access is the killer application.' *Journal of Academic Librarianship* 22(5), 1996, pp.371-375.

**Subscriptions**

*Aslib Proceedings* is published ten times per annum and supplied to corporate members of Aslib. For 1999 the annual subscription rates are £170/$280 (£136/$224 to members of Aslib). For subscriptions and back issues please contact: Portland Press Ltd, Commerce Way, Whitehall Industrial Estate, Colchester CO2 8HP, United Kingdom. Tel: +44 (0) 1206 796 351. Fax: +44 (0) 1206 799 331.

Aslib, The Association for Information Management, is a world class corporate membership organisation with over 2000 members in some 70 countries. Aslib actively promotes best practice in the management of information resources. It lobbies on all aspects of the management of, and legislation concerning, information at local, national and international levels. Aslib provides consultancy and information services, professional development training, conferences, specialist recruitment, Internet products, and publishes primary and secondary journals, conference proceedings, directories and monographs.

Further information can be obtained from: Aslib, The Association for Information Management, Staple Hall, Stone House Court, London EC3A 7PB, UK. Tel: +44 (0) 20 7903 0000. Fax: +44 (0) 20 7903 0011. Email: aslib@aslib.co.uk *http://www.aslib.co.uk/*

Periodicals Postage Paid at Rahway NJ and at additional mailing offices.
**POSTMASTER:** Send address changes to PROCEEDINGS, c/o Mercury International, 365 Blair Road, Avenel NJ 07001

©1999 Aslib and contributors ISSN 0001 253X

ISSN 0001 253X
Volume 51
Number 9
October 1999

25/2/2000
wel

# Aslib Proceedings

New information perspectives

The future of 'free' information in the Age of the Internet
David Lennon

Copyright and Internet fanzines
Charles Oppenheim and Margaret Turner

What do we know about public library use?
Ian M Smith

The net generation: the experiences, attitudes and behaviour of children using the Internet for their own purposes
Peter Williams

Aslib THE ASSOCIATION FOR INFORMATION MANAGEMENT

*Aslib Proceedings* is a long-standing professional journal of international repute, which has a large and wide subscriber base. The journal brings currency, authority and accessibility to the reporting of current research, issues and debates in the broad area of information work. Above all the journal wishes to provide research and comment in a form that is easily and quickly understood: a fresh, rigorous, but unfussy, writing style is what is aimed for. The needs of busy practitioners (and academics) are very much in mind. Articles are refereed to meet the need for accredited and authoritative information. To ensure that this does not result in an unacceptable loss of currency, articles are speedily refereed by the journal's large, eminent and multi-disciplinary Editorial Board (assembled for just such a purpose). In exceptional cases they may be referred to an external panel of consultants. The journal's monthly publishing pattern helps in maintaining the currency of reporting. In consequence the submission of research in progress articles is encouraged.

The boundaries of information work are rapidly changing and widening. *Aslib Proceedings* has a role to play in the repositioning of the information profession by encouraging submissions from information managers, providers, packagers and consumers of all kinds – and especially those from the related information fields of journalism, electronic publishing, communication and Internet studies. An interdisciplinary Editorial Board has been assembled to assist in the process.

Prospective authors should submit articles of 2,000-6,000 words in length. Articles should be accompanied by a 100-150 word informative abstract and the numerical form of referencing should be adopted. In style and tone, submissions should fall between that of the journal's sister publications: *Managing Information* and *Journal of Documentation*. Articles should be furnished by email or on disc to David Nicholas, Editor, Aslib Proceedings, Department of Information Science, City University, Northampton Square, London EC1V 0HB. Email: nicky@is.city.ac.uk. The Editor is happy to provide advice on an informal basis to potential authors.

**Editorial Board**
Dr David Nicholas, *Head of Department, Department of Information Science, City University;* John Akeroyd, *Director, Learning Resources, South Bank University;* Sarah Blair, *Head of Publications, Aslib;* Peter Chapman, *Information Manager, NewQuest North East;* Professor Peter Cole, *Head of Department, School of Journalism, University of Central Lancashire;* Dr Tom Dobrowolski, *Senior Lecturer, Institute of Information Science and the Book, University of Warsaw;* Harry East, *Senior Research Fellow, Department of Information Science, City University;* Mike Hyman, *Managing Director, Head Software International;* Robin Hunt, *Creative Director, AreHause;* Nat Lievesley, *Network Manager, Centre for Policy on Ageing;* Helen Martin, *Information Manager, Research and Information Unit, The Guardian*; Dr Ian Rowlands, *Lecturer, Department of Information Science, City University;* Richard Withey, *Director, New Media, The Independent;* Ralph Adam, *Freelance Journalist.*

**Editorial Team**
*Editor:* David Nicholas
*Deputy Editor:* Ian Rowlands
*Editorial Manager:* Kay Nicholas
*Editorial Assistants:* Peter Williams (City University),
Colette Coles (London Research Centre)

# CONTENTS

285-289 The future of 'free' information in the Age of the Internet
David Lennon

290-301 Copyright and Internet fanzines
Charles Oppenheim and Margaret Turner

302-314 What do we know about public library use?
Ian M Smith

315-322 The net generation: the experiences, attitudes and behaviour of children using the Internet for their own purposes
Peter Williams

# The future of 'free' information in the Age of the Internet[i]

*David Lennon*[ii]
*LEXIS-NEXIS EUROPE, International House, 1 St Katherine's Way, London E1 9UN*
*Email: David.Lennon@lexis-nexis.com*

**Abstract** *As we have developed through the ages, tools have always been the key to adding value to both individuals and society. Information and knowledge are the high-value tools of the present age. The rise of the Internet has led to a preoccupation with 'free' information services. This article looks at how 'free' information really is and whether we can continue to expect high-quality information to be available without cost.*

## The counterfeit coin

It has taken us many generations to evolve from the Stone Age, through the Bronze Age, the Industrial Age, the Post-Industrial Age, to where we are now: the Information Age – or, at least on our way to entering the Information Age via a protracted stay in the Information Technology Age.

During theses successive Ages, getting the best use of the tool of that particular Age spelt power and fortune. In the Stone Age, it was the ability to make and use flint tools for hunting and farming, which secured material advantage; in the Industrial Age, such benefits came from harnessing natural and man-aided processes and materials. This created individual and societal advance, thus bestowing the greatest benefits on the inventors and exploiters of these developments. The same applies to the Information/IT Age.

Information is not only the tool, but, crucially, also the currency of the Information Age: we are in serious danger of debasing this currency by not adequately recognising its value, or the value of the people who originate, elaborate and circulate it. 'Bad money drives out good' was the tenet of Sir Thomas Gresham, founder of the English Royal Exchange in the mid 1500s, and, if we substitute the currency of this new Age, we now read 'Bad information drives out good'. Anyone who has tried to tune in to a worthwhile television channel from their New York hotel room will have had direct experience of this, but there are far more serious implications in the business and political environment than in the entertainment sector.

In this article I shall echo André Gide in his 1925 book *Les Faux-Monnayeurs*[1], who appealed: 'Il faut donner un son pur, authentique' (one must emit a pure, authentic tone). Otherwise, he concluded, the counterfeit coin, as it passes from hand to hand, increasingly distinguishes itself from the genuine article, until the final recipient finds he has traded something of value for something which is worthless. This is the potential danger of taking information from an unknown source 'at face value'.

Commercial hosts aim to collect the most authoritative information available (the 'best thinking'), to put it into a context which enhances its value to those who require it for decision-making, and to recompense appropriately those organisations and individuals who originated it. Without this recompense system, it is clear that the information stream will sooner or later be uneconomic to produce, and will be substituted by inferior sources. As the information stock then deteriorates, so will the quality of decisions on which they are based. Bad information drives out good, with all the economic consequences.

## Information as currency

In this new Information Age, information circulates at the speed of thought: weightless, stripped of the medium on which it had been dependent since the invention of writing, of papyrus, and, later, of the output from Gutenberg's and Caxton's printing presses. It is the best use of information tools that represents the greatest single opportunity for personal and for societal advancement: it is no coincidence that the world's two richest individuals, Bill Gates and Warren Buffet – among many others – trade in information and the tools to enhance its utility. As in the early economies, which were barter-based, information is not only a tool, but it is also to a very large extent the currency of the Information Age.

We are already an Information Technology – an IT – economy. However, until now the emphasis has been not on the 'I', but on the 'T', not on information, but on technology – as was the case in the days of Gutenberg and Caxton, not the message, but the message-bearer. This, I believe, is destined to change.

Such a move, from technology to information, brings us two hypotheses. The first asks: in what way is information a currency – what are the common characteristics of money and information? These are threefold:

*Information can be traded or bartered.* Perhaps the most direct working example of this is the finance sector, where stockbroker research – which costs millions of dollars to produce – is 'given' to the investment community in order to stimulate trades executed by the donor institution – a clear instance of barter. In an increasingly service-oriented economy, while consultancy may not yet be bartered, 'information' and not some tangible goods is the root of their proposition, and has a clearly defined value in the eyes of the customer.

*Information is a 'store of value'.* Witness the huge information banks which exist around the globe, and on the sale of which whole industries have developed, from high-margin, time-critical financial information, to the vast archives of legal and regulatory information, or of scientific, technical and medical data and latterly the increasing trend towards 'marketing information' – i.e. personal facts about you and me, such as purchasing habits – so zealously collected by e-traders.

*Information can be invested in order to create more value.* One example of this can readily be seen at conferences and seminars, as delegates compare notes and deepen their own learning with that of their colleagues in order to increase their value to their organisation.

As Knowledge Management has been enriched by the vocabularies of finance and accounting, many expressions, such as 'leveraging corporate intellectual assets' or 'information audit' have been carried over. This further underlines the commonalties between information and liquid assets.

The second hypothesis asks: what makes me think that we are in the Information Technology Age and only now moving into the Information Age?

$1.8 trillion is now spent worldwide on information technology; this represents 6% of global GDP. It has long been predicted that hardware would be replaced by software as the more expensive item. This is now the case – and it is surely an ineluctable development that, as software development increasingly is either shipped offshore or automated to reduce costs, the information whose transport through the veins and arteries of the information of the economy is facilitated by software will increase in value. And those who – as in the Stone Age and the Industrial Age – use it to their, and their organisation's, best advantage, will be the beneficiaries of this development. They will also be the principal architects of the new organisational and societal structures which evolve as a result.

As manufacturing becomes better, cheaper, faster, more jobs are shifted from production to the provision of services to enable us to continue to work better, faster, cheaper, and information is at the heart of this evolution.

## Evolution and the knowledge worker

In the Industrial and early Post-Industrial Age, the principle of 'best practice' was widely espoused. In the Information Age commercial hosts can be seen as providers of 'best thinking' – the largest collection of the most relevant information in its most appropriate form for the customer, whose job it is in the Information Age – when information is not only the raw material but also increasingly the coin – to join Gide in his plea. This can be done by validating the information coming into our organisations and ensuring that the highest-quality information is available to enable us to make the best possible decisions.

Certainly this information has its price, but, in the long term, the price will be lower than would have resulted from reliance on a business model, as is so prevalent on the Web, which neither sees the originators (be they analyst, journalist, lawyer or whatever) sufficiently rewarded for their own inputs, nor intermediaries adequately rewarded and respected for their skills in obtaining and appropriately circulating them.

But perhaps there is some factor which means that the economics of production in the Internet Age differ from those in the previous Age – how else can the stratospheric values of 'Internet stocks' be justified, many of which have yet to turn a profit? Given that the costs of producing and investing in knowledge are increasing, why is this not yet reflected in some of the offerings available to us – particularly on the World Wide Web? Although I can now start an electronic newspaper on a shoestring (no printing presses, no paper costs, no legal department, no HR department, an electronic instead of a print archive), the day that I produce a defamatory article, or that I have a labour dispute, my e-customers will soon realise that the product and its credibility soon become vapourware.

## The cost of 'free' electronic products

The actual cost of a web site is $3.1m per annum[2] and rising rapidly. So why would a publisher want to give me this for nothing? Bad news – he does not. Because of the limitations of payment systems, the need for security considerations on the Internet, and the experimental nature of most electronic publishing initiatives, he is taking a calculated risk. Unfortunately, as he sees an increasing percentage of his print revenues migrate towards electronic, he will need to recoup his costs, because he still has substantially the same infrastructure to support. He can cover some of the cost by advertising, but customer tolerance is low, since advertising is more intrusive on the Web than on the printed page; the certainty is that the customer will pay.

One 'coin' that now has a higher value to the vendor in the information economy is bio-data: the vendor can track the items you are interested in and sell your personal details to vendors of other items he thinks may be of interest to you. While this may be acceptable in a consumer environment for books and CD's, if you are a business working in a competitive environment, you need the assurance that your business is your own and nobody else's.

## The free aperitif

To Milton Friedman 'there is no such thing as a free lunch'. Readers of *Aslib Proceedings* may have invested in an Aslib membership or subscription to this publication in the belief that paying the appropriate sum and 'losing' productivity by attending meetings and reading publications will benefit their organisation to, at least, an equivalent value. If Microsoft and Netscape/AOL are prepared to 'give' us their browsers today – and I would not even dare to begin to estimate the development costs – it is just possible that this is commercially motivated, either, as the American justice system is currently deliberating, in order to make life difficult for their less financially-secure competitors, and/or as an opening gambit in order to gain substantial market share in the exploding Internet market. Is it not likely that, once these objectives have been met, they will want to see a return on their investment, and that these expensive utilities will then become chargeable? There is no such thing as a free lunch, but it is not an uncommon opening gambit in certain restaurants to provide a free aperitif.

## Free information or 'fools' gold?

A customer requires stability in its relationships with suppliers; stability means infrastructure, and infrastructure costs money. Internet production and connectivity may have saved on the physical production costs, but the others remain the same. My own organisation, LEXIS-NEXIS, like many other hosts, is a commercial organisation which operates to return a profit to its shareholders and partners. This is the customer's assurance that we aim to be working with customers to our mutual benefit for as long as they are convinced by our value proposition. The situation, where information is being produced at a cost and provided free of charge, is not sustainable, so, while we should take advantage of this apparent generosity in the short term, we should recognise that it might ultimately prove to be ‘fool’s gold’.

So, if information is the coin of the new economy and we are no longer production workers, but knowledge workers, can we draw up a balance sheet of the actual costs of the production, distribution, and utilisation of this, the coin and the primary raw material of the Internet economy?

In addition to the infrastructure and associated maintenance costs, there is the very obvious cost of losing staff in cyberspace who, along with their valid Internet research, may also surf off into domains of personal interest – some of which can be very demanding not only in terms of time, but also of bandwidth. There is also the increased risk (and associated cost) of importing a virus into the organisation.

On the Web, many things are not what they seem – just as people can assume fictitious identities to hide behind their email activities, cyberspace prevents us from distinguishing between credible, authoritative publishers and ‘advertiser-sponsored’ commercials. And perhaps the greatest cost of all is the cost of misinformation. Since information is now distributed at cyberspeed within an organisation, this ‘false coin’, once injected into a company’s corporate consciousness, can be very difficult to remove: acting on it could have very expensive consequences for an organisation.

## The balance of probabilities

If one assumes that for each of the 24,000 publications contained in the LEXIS-NEXIS database, there is an average of 100 journalists, analysts, lawyers, commentators. That makes a reservoir of approximately 2,400,000 dedicated thinkers and writers whose authoritative and unbiased information is available to you. That represents an annual wage bill of, say, $240 billion. The balance of probabilities is that this particular collection of best thinking will cover most of the information needs of its target audience, and certainly more than any individual organisation could afford to support independently.

The Web is, then, running a considerable deficit in the sphere of business-to-business information; however, that is certainly not to say that all information on the Web is without value (academic, scientific and technical circles being a clear exception). This excludes fields such as patent and other information which is provided free by government organisations, since, while free to the consumer, the actual bill is picked up by the taxpayer.

## The need for due diligence

It cannot be a positive sign for the information industry or its customers that two global giants and competitors in the financial information sector – Reuters and Dow Jones – decide to combine their aggregation businesses in order to enhance their chances of commercial success. Reuters’ fortunes particularly were built on providing timely information, starting with carrier pigeons between European cities in 1849. And even this proposed concentration within the information industry is only the latest of mergers, or even failures. There is no such thing as a free lunch: if we as customers do not insist on commercially viable models, the certain consequence is a reduction in the choice of our suppliers as vendors such as Data-Star, Dialog (in its initial form), Apple e-world, Europe Online, Maxwell Online, Individual, CompuServe, Genie and many others which have gone out of business, merged, or else changed their business models in order to secure their future.

So let us, as knowledge workers at the dawn of the Information Age, and therefore its principal architects and beneficiaries, not accept Gide’s pessimistic conclusion to *Les Faux-Monnayeurs*: ‘Dans la vie, rien ne se résout, tout continue’ (‘In life, nothing is resolved, everything continues’) but fulfil our function as custodians of our organisations’ information assets by exercising due diligence over its quality as it is introduced, and by ceaselessly questioning its actual value and ultimate cost. While it might appear to be free in the short term, we might find the long-term cost not to be justified when we uncover the full facts.

## References

1. Gide, A. *Les Faux-Monnayeurs*. (Paris; Gallimard, 1925.)

## Note

i. Translated and revised version of paper presented to the GFII (Groupement Francais de l'Industrie de l'Information) conference on the Impact of 'free' Internet services on fee-based services (L'impact de la 'gratuit' d'Internet sur les services payant d'information), held at Maison de l'Europe, Paris on 27th May 1999.

ii. David Lennon is Director, Sales and Marketing, with LEXIS-NEXIS EUROPE in London, but the views expressed here are his own and do not necessarily reflect those of the Company.

# Copyright and Internet fanzines

*Charles Oppenheim and Margaret Turner*
*Department of Information Science, Loughborough University, Loughborough, Leics. LE11 3TU*
*Email: C.Oppenheim@lboro.ac.uk*

**Abstract** *An investigation of what happens when fans take material related to popular TV science fiction/fantasy programmes without permission, give the copyright holder full credit, and use it to promote the show from which they have taken the material. The web sites relating to two TV series were chosen for study: Forever Knight and the X-Files. A large number of owners of relevant fan sites were surveyed by questionnaire. The investigation showed that the approaches taken by the copyright holders in the two case studies differed quite markedly. Fox (X files) takes a harder line than Sony (Forever Knight) appear to have done. The issue of rights ownership is becoming increasingly important, especially with the advent of the Internet. Though considered by many to be something of a trivial area, fan activities have in fact been at the forefront of the copyright conflict taking place on the Web.*

## Introduction

For centuries, intellectual property laws have protected the interests of writers, performers, artists and others. However, recently, with the wide variety of possible media and the widespread availability of duplicating technologies, policing these laws has become increasingly difficult. Nowhere is this more evident than with the World Wide Web (WWW). Freely accessible to millions across the world, the Web has become a source of controversy in many areas. Issues of privacy, censorship and, of course, copyright has been debated almost since its inception, but for none of these issues has a satisfactory resolution yet been found.

The popularity of the WWW is largely due to the fact that, through it, anyone can publish anything and potentially have an audience of millions. Unfortunately, while original works are common, so too are illegal copies. If someone duplicates the entire, or a substantial part of a text and puts that on the WWW, they are infringing copyright. In this paper, we consider what happens if someone takes material related to a popular TV programme without permission, gives the copyright holder full credit, and uses it to promote the show from which they have taken the material.

Online fans differ from their off-line counterparts in that it is much easier for them to express their enthusiasm for their subject, via mailing lists, chatrooms, newsgroups and, of course, WWW sites. This is particularly true of fans of the science fiction/fantasy genre, who are already predisposed towards technology. Added to this is the fact that, in the case of television series, a wide variety of audiovisual material is broadcast into their homes on a regular basis. With the combination of easily accessible source material, ready availability of duplicating technology, and the creative enthusiasm of the fans, it is not surprising that screen grabs, sound clips and even short movie files find their way onto the WWW. Visitors to that web site can then easily take copies of the material for personal use or for use on their own web site.

Such cross-pollination of materials for web sites is very common. However, the copyright holders, usually large multinational companies, are becoming aware of the existence of these sites and their unauthorised use of copyrighted material. Naturally, the rights holders want to protect their property whilst fans want to be able to share their enthusiasm for a show. Potential conflict is the inevitable result.

## Background on the TV series

For this research, we decided to examine the web sites relating to two TV series. These are briefly described below.

### *Forever Knight*

*Forever Knight* was a science fiction/fantasy series with Gothic overtones; depending on individual perspective it could either be regarded as a series about a policeman who also happened to be a vampire, or a vampire who also happened to be a policeman. The series focused mainly on themes of faith and redemption and was considered to be one of the more thought-provoking series in circulation at the time.

*Forever Knight* began its life as an American made-for-TV movie 'Nick Knight' in 1989[1]. The movie was a failure but inspired some interest in Canada. In 1992, the pilot was remade in Canada, and was renamed 'Dark Knight' with some minor plot alterations and almost entirely recast using Canadian and British actors. This remake succeeded, spawning a series, which ran from 1992 to 1996.

The series struggled throughout much of its existence. At one stage, significant changes were made to the series to make it more mainstream; this involved several cast changes and a shift in the overall tone of the show. It appears that these changes succeeded in little more than alienating a portion of its original audience and it is possible that this was the cause of the show's eventual demise.

Despite its short existence, the series was aired in a number of countries including Canada, USA, UK, Australia, New Zealand, Germany and South Africa. It has also produced a few merchandising items, namely three novels, two of which were written by fans, and two soundtrack CDs. Significantly, several items appeared after the show's cancellation; most recently the second CD was produced in December 1998 after a concerted effort by the fans. Much of the success of these fan efforts is due to their online community, focused on the ForKni-L mailing list with approximately 1,000 subscribers[2]. Fans have even managed to keep the series in some sort of life through 'virtual' materials written by fans and posted to the FKFic-L mailing list. This 'virtual' series reflects what the fans would want to be series 4 had it lasted that long, and essentially tries to remain consistent with the series. This is in contrast to normal fanfiction, which is usually biased towards the author's preferences and is not obliged to stick with the commonly accepted views of the series. A number of the fan sites contain information on how to join these mailing lists. It is perhaps apt that a series, in which several of the main characters are undead, should be so unwilling to yield to the inevitable.

### *The X-Files*

The *X-Files* is a well-known science fiction series; originally based on real-life mysteries involving possible paranormal phenomena. It later focused more on the realm of conspiracy theory and UFO's. Initially intended to be something of a cult series, it very quickly reached mainstream appeal and has since resulted in its own spin-off series, called *Millennium*.

The series began in late 1993[3] and quickly attained a large worldwide audience. Financed and controlled by Fox Networks and its producer, Chris Carter, the series has continued to grow in popularity for some time, although interest now seems to be declining somewhat. As might be expected with such a large corporation in charge, once its surprise success was revealed, a great deal of effort went into promoting it further. There have been spin-offs, such as at least one computer game, and an array of other merchandising including mugs, clothing, key rings, books, posters and other paraphernalia. Its success can also be seen in the way that the series' motto 'The Truth Is Out There' has entered common usage.

The *X-Files* is currently showing in virtually every developed country worldwide, so it is unsurprising that its fan base is also multinational and therefore finds the WWW to be a most effective method of communication between fans. However, probably because of its size, it is not so well organised or concentrated as the *Forever Knight* fan base. Further dispersion across the WWW appears to be partly due to the actions of Fox in tracking *X-Files* sites for copyright infringement. As their crusade continues, the sites become harder to find. In fact, typical search tactics for the WWW do not yield great success and searchers are forced to use site to site linking to locate sites. Although Fox's actions have caused some considerable consternation, there is no real organised attempt to counteract it. There are a vast number of mailing lists, newsgroups and chat rooms for the *X-Files*, but they are quite dispersed and individually do not have very high membership.

It is ironic that in trying to shut down *X-Files* sites in order to maintain full control over its product, Fox is behaving much like the 'big brother' organisation portrayed so negatively in the series.

## The fan sites

Fan web sites are created largely to share a fan's passion for a particular series or several series. The fans try to include things that will give the visitor something of the flavour of the show. The fans often use all the media available to them to get across their message. The overall scale of the phenomenon is impossible to accurately measure, as new sites go up, old ones come down and others change location. However, it is probably a safe guess that there are about 100 *Forever Knight* sites and possibly ten times that for the *X-Files*.

Some types of media are more popular than others, partly because of the difficulty in creating certain types and partly just due to the fans' preferences.

### *Sound*

This includes .wav (wave), .mid (midi) and .ra (real audio) files. Wave files are the simplest to create, and are the most common form of sound file. They vary in quality depending on the creator's decision between file size and sound quality. They are a direct copy of sound from the show including speech, background noise and music. Midi files are not unusual, but require much more effort to create as they are made by programming a tune into the appropriate software. This requires some ability on the part of the fan and while the tune is copied, the end file is a creation in its own right. Quality of playback depends on the sound card of the visitor's computer. Real audio files are quite rare in this particular area, though not uncommon on the WWW as a whole. The best parallel to draw is with radio broadcasts since real audio files cannot generally be downloaded to play offline. Quality and type of sound is much like radio, in that it varies widely depending on external circumstances. Curiously, .mp3 files – high compression files of almost CD quality that have caused considerable concern in the music industry – have yet to be used on TV fan web sites.

### *Stills*

These primarily refer to screen grabs or snapshots taken directly from the television screen. However, it can also include the scanning in of publicity stills and composite pictures created from screen grabs and stills. File formats can be .bmp (bitmaps), .gif (gifs), .jpg (jpegs) or other graphics formats. The quality varies with the skill of the creator and is generally good.

### *Movies*

These are usually .avi files and are not particularly easy to create. They are, in essence, a brief section of the show, complete with sound and moving images, exactly as it would appear on television. Quality can vary quite widely, again dependent on the skill of the creator and the technology used. The best of these files, while excellent in its usual size, when expanded to full-screen loses definition quite badly. The less well created files can become virtually useless if this is done, but are acceptable at their normal size.

### *Quotations*

These are usually short quotations from the show, but may extend into long excerpts or even full transcripts. Usually these are just gathered from watching the series, but may in some cases actually come from genuine copies of the scripts used in the series. Errors are not uncommon.

### *Logo*

This refers to the use of the series logo or a reasonable facsimile.

### *Fan-fiction*

This refers to fan-fiction, often referred to as fanfic. These are literary endeavours on the part of the fans that make use of characters, situations and concepts from the series and may include crossover stories between more than one series. This has existed for quite some time prior to its appearance on the WWW. Rebecca Tushnet has traced it back some way[4], but probably the point at which it became a staple of the fan world was with the advent of the original *Star Trek* series. Fanfic varies enormously in quality and style, from humorous to serious, from plot-driven to character development stories, and can include poetry or limericks. It is probably also worth pointing out that since fanfic writers are not constrained by what can or cannot be shown on television, fanfic often can be of an adult nature in terms of language, violence and sex. For example, it can involve homosexual relationships between series' characters. Many archives exist on the WWW for fanfic and there is often a central repository, although there may be many other smaller archives. Most fan-fiction makes its first appearance on mailing lists, and only later is posted to the WWW, sometimes in plain ASCII text, sometimes in HTML with pictures and the like added.

### *Other*

This is mostly merchandise, both official and unofficial, but also includes such things as small freely available games written by fans.

We identified over 100 relevant web sites using standard search engines. We noted details of the contents of these web sites. The overall results of the analysis are shown in Table 1 below.

Obviously, stills being the simplest, appear on the majority of sites. Some of the more specialised content types, like .avi files, which require greater technological knowledge, and fan-fiction, which requires literary inclination, are less common.

*Table 1. Analysis of fan web sites' content*

| Sounds | Stills | Movies | Quotes | Logo | Fiction | Other |
|---|---|---|---|---|---|---|
| 50.5% | 99% | 19.2% | 25.3% | 36.4% | 42.4% | 17.3% |

## Copyright and web fan activities

The electronic medium is a dangerous one for publishing. While in theory simpler to do and having a far greater audience globally, it is also one that lends itself to copyright infringement. Anyone can copy an image, some text, or a sound file to create something identical to the original. There is nothing to prevent it being re-used again on another web page without the copyright holder's knowledge or permission. If a fan wishes to create a site for a series, then it is natural that they would want to include material from the series on it. This material can be obtained two ways. It can be taken from an already existing site, or it can be created by the fan.

On top of this ease of duplication, there is also the problem of the enforcement of intellectual property laws. Many of these problems are a direct result of the very nature of the WWW and are seen by many as actually being amongst its advantages.

The WWW is, by definition, a global phenomenon and as such it cannot be said to fall into the jurisdiction of any single country or even group of countries. This causes difficulties when it comes to the legalities; firstly, under which country's legislation can a web site be sued, the country of its origin or the country in which the infringement was discovered? For example, a Korean *X-Files* web site containing screen grabs and sound files from the series, would it be taken to a Korean court or would it be done in the US as that is where the copyright holders, Fox, are based?

In a similar vein, the sheer size and complexity of the WWW means that policing such a constantly changing environment is, for all intents and purposes, an impossible task. This is compounded by the level of anonymity the WWW provides. Most web site designers like to be known for their work, but it is not unusual for a web site to provide no indication whatsoever either of the person behind its creation or the country of origin. Although some of the information can be tracked through the ISPs, this is not always an option.

There is a defence against a copyright infringement action known as 'fair dealing' in the UK, or 'fair use' in the USA. This defence means that a user may make copies of copyright materials so long as the copying is for one of the permitted purposes, and as long as the copying does not damage the legitimate interests of the copyright owner. Fuller details about the defence can be found in standard copyright works[5]. Unfortunately, the permitted purposes vary from country to country, but the most common such purposes are: research; private study; teaching; reporting current events; criticism or review; or for purely private (home) purposes.

Whilst one can argue that a private individual downloading certain fan materials for their own private enjoyment may well be fair dealing or fair use, the placing of such materials on a web site accessible to all will certainly never be fair dealing or fair use.

Though the law is clear that copying is copying regardless of the medium, it has to be acknowledged that the Internet is a difficult medium to impose this law on. Its regulation is considered by some to be well nigh impossible, while others feel that it should not be regulated on principle, quite aside from the logistical problems. As such, the conflict lies across two issues: what is or should be considered illegal copying in the Internet environment; and the difficulties of regulating such a fluid medium that transcends so many conventional boundaries.

## How have the rights holders reacted?

### *The X-Files*

The *X-Files* series is owned by Fox Networks. While there are no reliable figures as to how many fan sites have been forced to shut down by Fox's actions, the shut downs appear to have become something of a matter of course over the last three years. Fox's tactic appears to be simply to identify a site and its creator, and then to issue them with a warning letter, threatening legal action if the infringing material is not removed. As yet, none of these cases have made it to court, with most sites shutting down, though some have appeared elsewhere on the WWW.

In response to protests over the closure of sites, Fox issued a statement on their web site as follows:

> 'In response to erroneous rumors circulating on some internet web sites, Fox wishes to assure fans of "THE X-FILES", "MILLENNIUM" "KING OF THE HILL" and "THE SIMPSONS" that it is not Fox's intention to shut down bona fide web sites created by fans devoted to these programs.
>
> However, Fox must require all web sites using copyrighted and trademarked materials to comply with guidelines that protect the creative integrity of the programs they represent. Fox has not asked any web sites to be closed except a few which were using program materials in an inappropriate manner. Some examples of inappropriate uses we have found include:
>
> (a) using copyrighted photographs, images and sounds which have not been approved for distribution;
> (b) removing Fox's copyright and trademark notices and asserting ownership by a third party, or asserting that copyrighted and trademarked materials are in the public domain;
> (c) copying or altering text, photographs, images, designs and logos from official web sites without permission or attribution;
> (d) copying or altering photographs, images, designs and logos from programs in a manner that denigrates the programs, the actors or the characters they portray;
> (e) using photographs, images, designs and logos to promote a product or commercial business or service;
> (f) carrying advertising for parties not associated with the programs;
> and
> (g) offering unlicensed merchandise for sale;
>
> We know true fans do not encourage inappropriate activities and respect the hard work of the many talented and imaginative people who make Fox's programs unique. Fox appreciates the support of all fans in protecting its creative rights and proprietary interests.'[6]

This approach appears to have developed with Fox's awareness of the WWW as a marketing tool, since it is reported that their initial response to fan sites was quite positive[7]. Fox are attempting to maintain a strict control over the *X-Files* in all its permutations, and see fan efforts as competition. The attitudes of the stars of the show are unknown, but it is unlikely that they would be as concerned about the online fan phenomenon, largely due to their lesser financial stake in the marketing and that, as actors, virtually any publicity is good publicity.

The first web site was shut down in early 1997 and Fox has maintained its stance over the infringements since.

### *Forever Knight*

*Forever Knight* is currently the property of Sony; it was created by James Parriot and Barney Cohen, and its main stars were Geraint Wyn Davies, Catherine Disher and Nigel Bennett. These parties are the main stakeholders and it should be noted that some of them maintain quite close contact with the online fans, either directly via the mailing lists, or through fans who are considered trusted intermediaries.

Unlike Fox, Sony has issued no formal statement regarding fan web sites and has given no indication that the existence of these sites concerns them at all. It would appear that Sony have quite a tolerant approach to the fans and are willing, if not to make a new series, then to take advantage of other marketing possibilities.

Amongst the available merchandise for *Forever Knight* are a number of books, at least two of which were written by fans who applied to Sony for licence to do so. A similar situation applied to the production of the second soundtrack CD which went on sale in the first week of February 1999. This was a direct result of fan efforts on the WWW with the cooperation of Sony after sufficient assurances were made that a certain number of sales would be made.

It appears that Sony are willing to overlook fan activities on the WWW and are not averse to taking advantage of a guaranteed market for certain merchandise.

Other stakeholders appear to be a little more wary of the power of the online community, probably because they have a higher level of personal involvement with the series. As a medium for communication and 'virtual' communities it is appreciated, but there is a strong concern regarding the possibility of misinformation and the unfortunate habit of rumours becoming believed if told often enough. This is a valid concern, and mailing lists try to limit this by stating that only recognised spokespeople are allowed to make announcements regarding stars and other 'officials'. However, this in no way prevents the spreading of rumours via private email or elsewhere outside of the jurisdiction of the mailing list rules.

There is a tolerance for many of the infringements, since they are seen as essentially harmless, 'but it is generally best to make such policies known from the beginning rather than trying to introduce them at a later date'. This was the opinion of one of the show's producers, who wanted to remain anonymous – personal communication 31 October 1998.

While there may be some definite concerns regarding the online fan activities, Sony appears to be largely tolerant, though whether this is the cause or the effect of the fans, obvious willingness to 'put their money where their mouth is', remains unknown.

We attempted to contact both Fox and Sony to assess their views on the issues raised in this article, but both declined to respond.

## Fans' opinions

As part of this research, we contacted a large number of owners of relevant fan sites with an email questionnaire. In this section, we consider the responses of the fans. For reasons of confidentiality, we have not identified any of the respondents (except where they were willing to be identified).

*Are you aware of the Trademark, Copyright and Moral Rights laws relating to the use of third party materials in the creation of your web site?*

### *Forever Knight*

The majority (69.3%) claimed some knowledge of the law, but only in a vague sense. There appears to be an acknowledgement of the general illegality of their actions. However, the specifics of the laws and penalties were largely unknown. Nearly a quarter of respondents (23.1%) had a more in depth awareness of the legalities involved, largely as a result of their profession, such as legal secretary, or education. This increased awareness does not necessarily correspond with greater abidance with those laws; for example, a legal secretary maintains a site that specialises in screen grabs and is quite willing to make and send out specific shots on request.

### *The X-Files*

In this instance 33.3% of respondents had a good knowledge of intellectual property law, while a further 53.3% had only an average awareness of it. This is quite possibly due to the greater legal activity by Fox. Again there is no correspondence between the level of legal awareness and abidance with those laws. One respondent with good knowledge claimed it was 'a rebellion'. Fox's action does tend to increase the likelihood of the presence of disclaimers on the sites, though.

*Have you contacted, or been contacted by the rights owners with regard to your web site?*

***Forever Knight***
No-one had been approached by any of the *Forever Knight* rights owners, although one fan had contacted two of the rights owners for permission to put sound samples and lyrics from the CD on his site and was granted permission to do so. There also seems to be some contact between the fans regarding use of each other's files on their sites even though they are not the actual rights owners. This seems to indicate greater comfort with approaching a friendly individual rather than a large corporation. Generally, people agreed with the view that '[the fans] are fortunate that the *Forever Knight* rights holders have been most generous in allowing FK fanfic to run rampant'.

***The X-Files***
Some (3.3%) respondents had been contacted by Fox in regards to their web site. Those contacted by Fox for the most part, no longer have a web site. One individual presented an interesting case study:

> 'I put my site up the next day after the *X-Files* premiere. I knew that the show would be a hit. I started by making screen shots with a digital camera and some alien AniGIFS. The next week, I was contacted by 1013 Productions with.... *a thank you note* that I was promoting the show. I of course, was pleased. I was the first *X-Files* web site up. It was small but I did get popularity going. Three years later, when the show was a hit, I got contacted again. They were no longer thankful. I was breaking copyright. I was younger then and I got scared. I was a new born publisher and didn't want this on my record. I took off the site for three weeks. I got petitions started....and my fans sent letters to all execs saying that what they did was cruel. I moved my site to [ISP deleted] and then got followed. Finally, I moved it to [ISP deleted] in 1998 when I was running out of room. They haven't contacted me yet.'

This appears to be the approach taken by a number of fans. The anonymity of the WWW makes it very difficult to track such activities.

*If the rights owners asked you to remove all copyrighted material from your site, what would your reaction be?*

***Forever Knight***
In response to this, no-one said they would not remove the material. However, 72.9% of those would display reluctance to do so and would try to reach a compromise or argue their case before acceding to the request. Some maintained that they would be 'devastated' or at the least, unhappy, at the loss of their work, but would nevertheless respect the rights owners' position.

***The X-Files***
The responses to this question differed quite strongly from those of *Forever Knight* fans. 26.7% of respondents said they would remove the material without complaint. 46.7% of fans would reluctantly comply with the request, but would be 'outraged' or 'very upset' and would protest, petition or otherwise object to their treatment. At least one respondent replied that they would seek legal advice on the matter. The main difference is that 26.7% of *X-Files* fans would *not* actually remove the infringing materials from the WWW, but would simply move the site to avoid legal action, meaning that Fox would have to start from scratch again.

*What is your opinion regarding copyright on the WWW? For example, do you think copyright should be respected and enforced, or do you feel that allowances should be made for copying that, in general, does no harm?*

***Forever Knight***
All respondents expressed dissatisfaction with the current legal situation. Many felt that fan

activities that make no profit should qualify for fair use, or that a specific exception should be made, since many consider their sites to be a form of free publicity for the series. A few felt that simply clarifying the current laws would help a great deal. One respondent felt that, in principle, it was fair of copyright holders to demand the removal of things already officially available for sale through them. She also felt that, in practice, this was not something that could easily be monitored or controlled. The general consensus seems to be that copyright holders are likely to lose more money employing lawyers and people to hunt down infringing sites than any theoretical loss of profit because of the sites themselves. No-one appeared to have any problems with acknowledging the proper ownership of the works, and many already include such disclaimers on their sites or with their works. Typical examples include:

> 'Disclaimer: Forever Knight is the sole property of somebody else, probably Sony/Tristar or somebody like that, but certainly not me. This page is not intended to violate any copyright laws. In fact, I encourage everyone to buy 20 copies of the Forever Knight Soundtracks and of each of the novels to make certain that "the powers that be" know that we fans exist!'[8]
>
> 'Disclaimers: All characters herein portrayed are the property of their Sony/Tristar and James Parriott. This fiction, and the use of these characters, is for limited circulation only and no profit is being made from their use. Any opinions or emotions exhibited by these characters is of my own invention only and is neither attributable nor necessarily intended by their creator/owners. Gosh, isn't that all nice and legally sounding?'[9]

### *The X-Files*

A number (6.7%) of respondents felt that current law was adequate for this situation, but a large majority disagreed with this view with varying degrees of vehemence, again hinging on the issues of profit and free publicity. These respondents felt that attacking fan sites upset fans and caused Fox to spend money pursuing a course of action that alienated fans, and possibly reduced viewing figures.

### Other Comments

The majority of sites are, unsurprisingly, American in origin; other English-speaking countries such as Britain, Canada and Australia represent most of the non-American sites, although French, German and Indian sites also exist, among others.

The issue of copyright is a particularly sensitive subject amongst *X-Files* fans. Perhaps not surprisingly for fans of a series that espouses paranoia with regard to faceless corporate entities; the arrival of the questionnaires occasionally led to fears of having been found by Fox and in one case the response was of outright hostility. We received many requests for reassurances that the identities of respondents and the location of their sites not be divulged. Once we gave these, they were happy to reply to the questions.

## Discussion

The approaches taken by the copyright holders in the two case studies differ quite markedly. Fox is taking a far more hard line approach than Sony appear to have done. Sony have yet to comment on the *Forever Knight* fan activities on the Web, although their knowledge of their existence is certain. Fan efforts for the series have resulted in more or less continuous airplay since the series' demise, both in the UK and the US, and have produced a second soundtrack CD, including newly recorded material, three years after the series ended. Although the chances of the series being revived at this late date are growing increasingly slim, the fans are still campaigning for more. They realised that, while producers and actors may hold some affection for the series, the financing companies need to be convinced in financial terms. To this end, Internet auctions, charity events, donations and even blood drives have been organised over the years, raising a phenomenal amount of money for charities chosen by the actors of the series. Surprisingly, the series is still gaining a few new fans as a result of its visibility on television and on the Internet and copies of the books and CDs are still being sold. So, without any real effort on the part of the copyright owners, they are receiving a relatively small but steady stream of letters and income. It is very likely that without the efforts of the fans both in cyberspace and out, *Forever Knight* would sink without a trace.

Fox have acted against a number of *X-Files* sites on the Web. The series is still immensely popular, but there is evidence of increasing dissatisfaction. In particular, despite its continued popularity, the fans have become increasingly vocal in protesting against Fox's actions. News travels very quickly on the Internet, and when the first site was threatened the deluge of protest emails crashed the server of the lawyer who served the notice. Fans' opinions of Fox have been lowered as a direct result of their treatment of the fans on the Web. Petitions and protest sites are fairly common and fans are more than happy to tell anyone their opinion of the situation. It seems likely that the *X-Files* will continue for a few years yet, but it seems that the number of fans, who would try to save it when it does go, is steadily shrinking.

Sony's approach has kept the fans happy, but one has to wonder if they would have done the same had the series not ended. Fox, on the other hand, in attempting to protect their rights to the series, appear only to have succeeded in creating considerable bad feeling amongst the fans. The number of *X-Files* sites has not decreased and neither has the content of those sites changed to avoid infringement.

Legislating the Internet is a very complex issue, simply because of the nature of the online environment. The legal issues involved are well explained in Gringras[10]. In addition, the plethora of parties involved, ranging from corporations to the public, and the sheer logistical difficulties of regulating such a fluid medium cause problems. A law that is completely impractical in implementation accomplishes nothing. Perhaps the most important consideration in the creation of any law is that of the ability to enforce it. This is probably the greatest hurdle facing anyone wishing to legislate the Web. By definition, the Web is a worldwide phenomenon. It crosses all the jurisdictional boundaries and effectively blurs the international borders to the point where they cease to have any real meaning. Web surfers have very little regard for the country of origin of the web sites they visit or download from and are even less likely to be aware of copyright laws particular to that country.

The removal of one version of a site does not require the removal of its mirrors in other countries and so they need to be sued individually. Does it make a difference if they are independent mirrors and were authorised, but not originated by the original web site creator? What liability does the ISP have in any legal action involving a web site maintained on its server? These sorts of questions have to be answered definitively if there is going to be any sort of successful attempt to control the situation.

A related issue is one of inconsistent copyright laws between nations. Many countries follow the same basic principles. However, there are some notable exceptions where the copyright law differs from that elsewhere, and in some cases is virtually non-existent.

Another issue is the apparent anonymity of the Internet, which is deceptive to a certain degree. There are ways of getting the information required about a web site or an email address, but it is equally possible to work around this. False credentials are the simplest and most obvious. This is an area where the status quo is always shifting as new technologies and hacks are developed. The anonymity that is currently available is sufficient to prevent casual identification and may be used to evade more in depth searches. If a company is unable to trace a site owner, they cannot sue them. The threat of litigation may persuade an ISP to remove a site even against the wishes of the site owner, but increasingly people are setting up their own servers. The further fragmentation of the Web allows greater freedom and flexibility for web site creators since they do not have to abide by the rules or limitation of the ISPs, but it also substantially reduces the level of control it is possible to attain over the Internet as a whole.

The identical nature of computer generated copies adds another layer of complications to the problem. The fact that anyone can save practically anything from a web site for personal use or for use on their own web site leads to a number of problems. The creator is not necessarily aware of downloads made from their site and have little control over this activity. It is generally considered good manners to obtain permission from the author, but this is not an especially widespread practise. Some surfers do have a tendency to consider the Web to be the literal expression of 'public domain' and will take anything without regard for ownership and do whatever they like with it. Generally however, these individuals are not particularly common within the fans, who seem to have developed a fairly friendly approach to such things. Financial recompense is unheard of, but there is simply a request to give credit where it is due, whether this is by ensuring the author's name is attached to the fan-fiction or by putting a small notice next to a screen grab saying who made it. In the face of such easy duplication of material, further measures of control or prevention are unlikely to achieve much success.

From the fans' point of view, the argument really is based upon the fact that the fan web sites make no profit from their use of the copied material. The copied material typically used by fans generally only consists of small extracts taken from the copyright material. Most importantly though, the fans do not attempt to sell the materials or pass them off as their own creations. Although they sometimes claim creation of a composite picture or fan-fiction derived from the TV series, since these involve personal interpretation and creative effort, most often they include disclaimers acknowledging the original copyright ownership of the characters and concepts. This does not excuse the action, but does indicate a willingness to 'play by the rules' as they perceive them. It also suggests that they do not see their actions as a form of theft, since they are quite clear about the origins of the material and are more than happy to point a visitor in the direction of that original source.

It is worth noting that the taking of small extracts, such as a single screen grab, is unlikely to be considered copying a 'substantial' amount of the original copyright material, and therefore is unlikely to be infringement. Nonetheless, if a copyright owner takes an aggressive stance regarding copying, a private individual is likely to be deterred from doing such arguably legal copying.

Many fans, in fact, look upon their web sites as promotional tools for their series, no different from television trailers or magazine adverts. Certainly the existence of these sites allows global advertising at no cost to the series' owners, the problem is that while some of the sites can be excellent, some can be very poor and to allow one is to allow the other. There is an issue of the connection of the series name with, for want of a better term, inferior products. This is very tricky and can only really be the decision of the company in question, although a number of sites do make a point of stating their lack of association with the actual owners. There have been some cases where exceptional web sites, namely the *Lurker's Guide to Babylon 5*[11], have been given tacit approval by the copyright holders in the form of early news and other things.

This sort of approach has gained the approval of fans because it allows for the provision of reliable information in a fan friendly format. Official sites tend to be very limited in content, often providing no more than an episode guide and viewing times for the fans who want something that will give a taste of the series itself. It is the fans' attempts to fill this perceived gap that results in the infringements of the series' copyright. Every TV series is looking for viewers because viewers mean money; fans are a dedicated form of viewer and as such mean more money – they won't just watch the series, they will buy the videos and the T-shirts and so on. However, a substantial fan base can be a double-edged sword; Henry Jenkins, a sociologist and researcher in the fan phenomenon, has identified a trend amongst fans towards greater personal involvement[12]. Fans lay claim to their own personal interpretations of the characters and concepts of the series; they love them and seem to feel that their loyalty to the series should give them a greater say in what happens to it. So while fans are a guaranteed source of income for the copyright owners, they expect consideration in return. They have their own views on the series and they feel they should have the right to express those views, whether it be general enthusiasm for the show in the form of sound files and pictures to attract new fans, or explaining their own perception of a character through fan-fiction.

To remove what the fans see as an entitlement is not something that is likely to go down well, as Fox have already seen. If they are prevented from getting involved with the series the way they want to be, they may lose interest in it altogether. The die-hard fans probably won't give up no matter what is done, but the majority of fans are more susceptible and losing them is to lose a fair proportion of the target market. The likely loss of income resulting from alienating the fans is something that copyright holders will have to weigh against the costs of policing the Internet for infringing sites and any legal actions taken.

Balancing all these issues is an extremely difficult task and suggests that, without considerable advances in the field of technology, it simply is not practical to legislate in this area at this time. If the problem is ignored by the various governments, they draw criticism and complaint from industry, but if they do act (such as with the EU Draft Directive on copyright) they draw fire from a fair proportion of the public.

## Implications for the LIS community

The issue of rights ownership is becoming increasingly important, especially with the advent of the Internet. Though considered by many to be something of a trivial area, fan activities have been at the forefront of the copyright conflict taking place on the Web. The web sites created by fans of TV series frequently use infringing material. The copyright holders want to maintain control of their product and retain their rights to exploit it. Where they see the fan web sites as a threat, some take action against them. Fans argue fair use and freedom of expression; copyright owners argue protection of their interests. There is no question that the law is on their side. The large corporations and governments hold all the advantages of resources and law, but the fans have the nature of the Internet environment to protect them. So far, neither side has won the conflict. The owners wield legal and financial power that the fans are unable to match, but when faced with litigation, many fans simply comply only to re-establish their site elsewhere on the Web, or it alienates them. Alienation is likely to reduce the success of any spin-off products, such as soundtrack CDs or computer games. In the end, no company can afford to ignore the financial power of their customers.

The governments seeking to regulate the Internet face the same problems of retaining control versus cost, both financial and in terms of consumer opinion. Technology, both hardware and software, is currently inadequate for the task; the costs in terms of manpower, telecommunications, technology and any legal actions will be high.

Despite the existing copyright laws, it may be that government and copyright holders alike will be forced to compromise with the fans in order to maintain some semblance of control over an otherwise unenforceable area of law. An example that could be used is that of the approach adopted by J Michael Straczynski, creator of *Babylon 5*, a popular science-fiction series. He has made himself available to fans via the Internet and maintains a presence on a number of the related mailing lists; the proviso is that fans refrain from certain activities, such as fan-fiction, while the series is still in production[11]. This appears to have had some success. For the most part, the fans are quite happy to accede to the request in return for his continued involvement with them. Obviously this approach is not going to be possible for all such situations, but it does offer some hope for a resolution to the copyright issues currently plaguing the Web.

## Possible solutions

The current law cannot be enforced and without that it simply becomes a waste of paper, respected by none of the parties. Indeed there is an argument that copyright is not an appropriate mechanism for protecting the owners' interests. A solution is possible, but it seems improbable without at least one of the three perhaps mutually incompatible conditions listed below, being met first:

1. Greater advances in computer technology may provide a solution to the otherwise

enormous costs involved in hunting down infringing web sites.

2. Copyright owners may attempt to establish a closer rapport with the fans and establish some sort of compromise agreement, ignoring what the law may state.
3. The Internet may become accepted as a medium beyond the control of anything less than a global ruling, and be acknowledged as a self-regulating nation in its own right.

This is an area where there are strong analogies to issues facing library and information managers in their relationships with copyright owners. There has been considerable tensions between the two parties, especially in regard to journal subscription prices. For many LIS managers, the Web offers a possible method of by passing the publishers, thereby allowing unrestricted use of materials without having to pay excessive fees. Although the situation is far better controlled than amongst TV series fans – we doubt that many librarians are routinely downloading materials illegally for their users to peruse – LIS managers should observe the debates going on amongst fans and copyright owners. The EU Draft Directive on Copyright[13], that could have such an enormous impact on the day to day activities of LIS staff, was primarily promoted by major media organisations rather than by publishers. It is clear, therefore, that LIS staff could suffer because of the types of tensions described in this paper. There is therefore a strong argument for LIS managers to keep a watching brief on the debates going on regarding fan activities. Whatever decisions are made are likely to have enormous repercussions for copyright holders, web users, and library and information managers. The approaches adopted by the multimedia and music industries may provide an indication of current and future trends in copyright law on the Internet.

## References

1. *Forever Knight. http://uk.imdb.com/Title?%22 Forever+Knight%22+*(1992) (visited 19 April 1999)

2. McDavid, L. Personal communication, 8 October 1998.

3. *The X-Files. http://uk.imdb.com/Title?0106179* (visited 19 April 1999)

4. Tushnet, R. 'Legal fictions: copyright, fan fiction, and a new common law', *Loyola of Los Angeles Entertainment Law Journal* 17, 1997, p.651.

5. Phillips, J., Wall, R. and Oppenheim, C. *The Aslib guide to copyright.* (London: Aslib, 1994-.)

6. *Ask Fox. http://www.fox.com/askfox.htm* (visited 19 October 1998)

7. Anonymous respondent. Personal communication (visited 31 October 1998)

8. *Welcome to Beck's Forever Knight Page. http://members.aol.com/BJDFKFan/* (visited 21 April 1999)

9. Schneider, M. Personal communication, 18 February 1999.

10. Gringras, C. *The laws of the Internet.* (London: Butterworths, 1997.)

11. Grimm, S. *The Lurker's Guide to Babylon 5. http://www.midwinter.com/lurk/lurker.html* (visited 18 November 1998)

12. Jenkins, H. *Textual poachers: television fans and participatory culture.* (USA: Routledge, 1992.)

13. European Commission. *Draft Directive on Copyright and related rights in the information society,* 1999. *http://europa.eu.int/comm/dg15/en/intprop/intprop/copy2.htm* (visited 14 June 1999)

# What do we know about public library use?

*Ian M. Smith*
*24 Courtenay Street, London SE11 5PQ*
*Email: iansmith@courtenay24.freeserve.co.uk*

**Abstract** *Data from Westminster Libraries on membership, borrowing frequency and stock use is compared with significant studies of public library use from the last 50 years. The article examines the extent of public library use, and the social and demographic factors influencing use. Analysis of book and audio-visual stock use indicates the concentration of demand on a few titles and subjects, and the relationship to what is in demand commercially. Aspects of public library use have remained constant over at least the last 50 years, and across geographical boundaries. Public libraries are used principally for leisure, for borrowing recent fiction, for useful non fiction related to immediate life interests, and for pop music. Quantitative data from library systems provides a different, probably more accurate, picture of the nature and extent of public library use.*

## Introduction

This article considers the nature and extent of public library use. It compares previous studies of public library use with data available from library records. Most earlier studies of public library use involved surveys, which can determine what kind of people use libraries, but interviewees habitually overstate their own library use. Comparing surveys with quantitative and computer generated library data permits a clearer view of the most valuable features of each. The results suggest that only a minority of people use libraries, that few of these do so regularly, and that the nature of library use has changed little over the last 50 to 100 years.

This study uses data from Westminster Libraries' computer records:

- membership records;
- library circulation data covering adult books, music and videos;
- data on frequency of use of members' tickets;
- a detailed sample of customer requests.

Westminster also collected personal information – age, gender, ethnic background – from those joining any of its libraries.

The study of a cultural activity such as public library use can never be an exact science, but the methodology assumes that similar results from a variety of studies and library data will lead to meaningful conclusions.

## Studies of public library use

The American *Public Library Inquiry* of the late 1940s reviewed all user studies from 1930 to 1949, carried out new survey work and analysed existing library records. It concluded that Americans did not use the public library service in the ways that the profession thought they did, and that only a minority of people used libraries.[1]

The inquiry's report on public library use established several themes, still supported by research 50 years later[2]:

- a small minority of adults regularly use public libraries;
- a small number of library users are responsible for most library borrowing;
- library use is determined by education;
- library users buy more books than non-users;
- most people do not use the library as a source of information;
- libraries are used for leisure, not to pursue enlightenment;
- library users have wider social and community interests than those who do not use libraries;
- most demand is for recently published material.

Studies of public library use in the 1960s and 1970s supported these findings. Groombridge used a mixture of survey work and analysis of library records to study public library use in London[3]. A comprehensive study based on Chester Libraries in the late 1960s, was considered by the author to have wide applicability:

> 'the striking thing, however, in the results, is the importance of universal fac-

tors like education, age and occupation rather than local differences'[4].

Luckham's work was followed by a study of library use in the London Borough of Hillingdon in 1972, incorporated into a study on the effectiveness of the library service[5].

A summary of 126 surveys of library use and reading from the period 1900 to 1976 reached similar conclusions to those of the *Public Library Inquiry*[6]:

- library use is related to education;
- reading is an end in itself;
- library users have wider leisure interests than non-users.

More recent research, exemplified by the Comedia report, *Borrowed time?*, relies heavily on survey data, probably overestimating library use[7]. Comedia emphasised the social role of the public library, and its impact on economic development. Conclusions were based mostly on interviews with interested parties, and focused public library studies on aspects of use that are difficult to measure quantitatively.

The Aslib *Review of the public library service*, commissioned by the Government, also relied entirely on survey data, rather than library records[8]. It also emphasised the social and economic impact of public libraries, but without any firm evidence, and without being able to measure that impact. Since 1991, Book Marketing Ltd has published an annual survey of library use, a source of valuable demographic data[9].

## How many users? – membership rates

Library membership has increased steadily since the nineteenth century, but may be reaching a plateau.

The Library Association and the Government both accept survey estimates that '58% of the adult population hold library membership'[12].

*Table 1. UK library membership 1924-1998*

| Year | UK population with library membership |
|---|---|
| 1924 (Kenyon)[10] | 11% |
| 1939 (McColvin)[11] | 19% |
| 1966 (Luckham) | 29% |
| 1998 (CIPFA) | 32% |

*Library membership records*

CIPFA now record the number of active borrowers in each authority[13].

*Table 2. Sample membership rates (CIPFA 1998)*

| Authority | Active borrowers as % of population |
|---|---|
| Camden | 28% |
| Lewisham | 11% |
| Barnet | 31% |
| Sutton | 38% |
| Richmond | 53% |
| Manchester | 26% |
| Liverpool | 19% |
| Sheffield | 22% |
| Metropolitan Districts | 30% |
| English Counties | 29% |
| **UK** | **32%** |

*Active borrowers = those who have borrowed in the previous 12 months.*

Nowhere does the membership rate reach the 58% claimed in surveys. *Books and the consumer* found that about 10% of adults claim to use libraries without borrowing books[14]. Some of these people may be library members. The figures in Table 2 do not distinguish library members resident in the area from those resident elsewhere, so probably overstate total library membership.

Westminster's membership database distinguishes resident and non-resident members, and deletes records if they have not been used for 12 months. Total membership is about 130,000, with an annual turnover of approximately 25%. One third of Westminster's members come from outside the borough. Despite a population increase of 13% from 1993 to 1998, the proportion who belong to a library has remained static, at about 40%. In 1999 there were 83,000 resident library members, about 39% of the population.

Westminster has a high ratio of library provision to population and a better-educated population than average[15]. It is likely that the documented membership of about 40% of the population is high compared to other areas.

Luckham investigated actual and claimed membership. He found that people typically overstated their library use, and checked the membership status of those who claimed to belong to the library; 46% of the sample claimed to be library members but only 31% actually were[16]. He concludes that people are often mistaken about library membership, and acknowledges that there could be a halo effect – interviewees say what they think the surveyor wants to hear. The Hillingdon study found that only 61% of those who claimed library membership had a current ticket[17].

The discrepancies found by Luckham and Totterdell reflect the difference between Westminster's actual membership rate of 42% and the commonly claimed one of 58%. The 1996 study of the Sheffield library strike found the same discrepancy without commenting upon it[18]. When surveyed people overstate their participation in a cultural activity such as library use, just as they exaggerate their voting in elections or likely visits to museums.

Both Luckham[19] and Groombridge[20] draw attention to lapsed members, indicating that 50% to 60% of the population join a library at some time in their life, but do not maintain continuous membership. Research in Enfield found that 48% of current non-users had used a library in the past[21]. Few people make continuous and frequent use of public libraries, but mostly use libraries occasionally, depending on personal circumstances and needs.

## Frequency of use

Westminster's database indicates when each membership was last used; these figures vary little from one quarter to the next (Table 3).

About a quarter of members borrow or return books at least once a month. Most members have not used the library in the previous six months. People use a library significantly less than they say they do. If more than half the library members have not used the library in the previous six months, and the annual membership turnover is 25%, this suggests a large amount of casual use. Regular monthly visits to change books are not how most people use libraries.

Library surveys conducted over a short time exaggerate the frequency of use, as the minority of frequent library users will form a disproportionately large part of the sample. Westminster's 1996 survey was conducted over one week and concluded that 80% of customers used the library at least twice a month[22]. The computer records, and the count of library visitors indicate that this is an overestimate.

## Frequent users and library issues

Library use, measured by loans, is concentrated on just a few users. Berelson examined the punch card records of Montclair Library, New Jersey, for 1947 and concluded that 'there is a heavy concentration of library use

*Table 3. Frequency of library use in Westminster*

| | % of adult members | % of child members | % of all members | % of resident population |
|---|---|---|---|---|
| Last month | 24 | 24 | 24 | 10 |
| Last 2 – 3 months | 10 | 14 | 10 | 5 |
| Last 4 – 6 months | 10 | 14 | 11 | 5 |
| Last 6 – 12 months | 56 | 47 | 55 | 23 |
| Within last 12 months | | | | 43 |

*Table 4. Distribution of issues and borrowers – Westminster*

| | Members | Visits/Year | Issues | Issues/Year | % of members | % of issues |
|---|---|---|---|---|---|---|
| Once a month or more | 33,000 | 42 | 2,092,500 | 63 | 24 | 86 |
| Every 2 to 6 months | 29,000 | 4 | 172,500 | 6 | 21 | 7 |
| Every 6 to 12 months | 75,000 | 1.5 | 169,500 | 2 | 55 | 7 |
| **Total** | **137,000** | **12** | **2,434,500** | **18** | **100** | **100** |

by a small group of frequent borrowers'[23]. Registered membership was 30% of the population, but nearly one third of these did not borrow anything in a year. He found that 15% of members accounted for 75% of library circulation.

In Westminster a minority of users – about 30,000 people – borrow books at least once a month. The majority of users borrow books only once or twice a year. In 1996/7 there were approximately 2.5 million issues and 2.4 million visits to lending libraries. Westminster's customer survey suggests that about 60% of those visiting a library borrow something, indicating about 1.6 million borrowing visits a year and 1.5 issues for each borrowing visit. In the customer survey about two thirds of those visiting monthly claimed to visit at least weekly[24]. With these assumptions it is possible to estimate the distribution of issues and borrowers (Table 4).

The findings are similar to Berelson's of 50 years earlier. *Books and the consumer* assert that regular borrowers borrow more books on each occasion than do less frequent borrowers, so it is possible that Table 4 underestimates the contribution of regular borrowers[25]. The small number of frequent borrowers account for most of the issues.

*Books and the consumer* has similar findings[26]

- 7% of borrowers account for 38% of issues;
- 18% of borrowers account for 62% of issues.

The Westminster figures suggest that only a quarter of library members – 10% of the population – account for 80% to 90% of borrowing. Research into individuals' borrowing patterns at Sandwell and Hounslow Libraries shows a similar pattern of use. At Hounslow Libraries 57% of members were light borrowers (13 items or less per year); 8% were heavy borrowers (more than 60 items per year)[27].

Library use is unevenly spread. The Library Association claims that '10 million people, drawn from all ages and social groups, use their library regularly, at least once a fortnight'[12]. The figure is probably nearer to 4 million people. If the Library Association estimates of use are correct, then annual library issues would be much greater than the current figure of approximately 500 million.

## Libraries and reading

The *Public Library Inquiry* established that if people read a lot they are more likely to use libraries. Those who own books, also use libraries more:

> 'The larger the home collection, the more use is made of the public library.'[28]

Groombridge also found that library members own more books than non-members.

> 'It is clear that the people who do in fact buy books in fairly large numbers and those who belong to public libraries are by and large the same people.'[29]

Luckham confirmed the relationship between book ownership and library use. Of those found to have no books in the home, only 6% belonged to a library[30]. Subsequent studies have confirmed that reading and book ownership relates strongly to library use[31].

Those people who read, use a variety of sources for their books and most studies have shown a roughly equal division between libraries, bookshops and borrowing from friends as the principal sources of books.

*Table 5. Source of reading – summary 1949 to 1992*

| Study | From a Library | Bought | Friends/ Gift |
|---|---|---|---|
| Berelson 1949 | 25 | 35 | 38 |
| Totterdell 1976 | 40 | 33 | 27 |
| Jackson and Martin 1991[32] | 33 | 39 | 27 |
| England and Sumsion 1995 | 30 | 41 | 28 |

*Percentage of adults currently reading. Figures may not total 100% due to differences in survey methods.*

Readers, book buyers and library users are overlapping, nearly coterminous groups. Librarians often assume that readers use different sources for different types of books. Sumsion assumes a distinction between books 'natural to own' and books 'natural to borrow'[33]. Analysis of public library demand (see below) suggests there is no evidence for this.

## Library use and social class

The Tottenham Libraries study established the link between social class and library use, and therefore with reading in general:

> 'Library members, like the book readers, are predominantly young, middle-class and well educated.'[34]

Every study links public library use with education and social class. The later the terminal education age, the more likely people are to use libraries. The *Public Library Inquiry* concluded:

> 'In every case differences in library use are much greater with regard to education levels than to any other factor.'[35]

All subsequent studies have confirmed this relationship between education and library use[36]. User studies in America 'have revealed over and over that education and income levels correlate with library use'[37].

Educated, middle class people make more use of libraries because they are better able to do so. Libraries do not change the social situation, they reinforce it. Those whom the library service was established to help the most, appear to make the least use of it.

## Library use and gender

The *Public Library Inquiry* found that more women than men are library members, and that women borrow more books than men do. All subsequent studies have confirmed this (Table 6).

*Table 6. Public library use by gender, 1949-1997*

| | % of Men | % of Women |
|---|---|---|
| Berelson 1949 | 41 | 59 |
| Taylor and Johnson 1973[38] | 43 | 57 |
| Stevens 1990[39] | 39 | 61 |
| England 1994[40] | 44 | 56 |
| Westminster Libraries 1997 | 42 | 58 |
| England 1997 | 43 | 57 |

*Children's reading choices* found that girls of all ages read more than boys, borrow more books than boys from public libraries, and borrow more frequently[41]. Table 6 does not reflect the distribution of borrowing between men and women. Women borrow more books than men do; heavy book borrowers are more likely to be women[42].

## Library use and age

The *Public Library Inquiry* calculated that those aged under 16 accounted for half of library members and 60% of library issues. Use was found to decline with each age level. Berelson recognised that this was directly related to education; in 1948 fewer older people had received a full education[43]. Westminster has a greater rate of membership among school age groups than among adults.

*Table 7. Westminster Libraries' membership by age (1995)*

| Age | % of residents who are members |
|---|---|
| 0-4 | 18 |
| 5-14 | 65 |
| 15-17 | 54 |
| 18+ | 41 |
| **All residents** | **42** |

In Westminster children also use the library more regularly than adults; 40% of resident child members having used the library in the previous three months, but only 34% of such adults having done so. Children aged five and over use the library even more frequently.

*Children's reading choices* found that the amount of reading and of library borrowing declines with age. Those aged 10 claimed to read more than those aged 14[44]. *Books and the consumer's* most recent research found that the percentage of those claiming to use public libraries was highest among 9 to 11 year olds, and then among 12 to 14 year olds[45].

The analysis of adults joining Westminster Libraries indicates that almost half are in their twenties; few older people are joining the library for the first time.

*Table 8. Age profile of Westminster Libraries' users*

| | % of adults using library (1996 survey) | % of new members (1997) |
|---|---|---|
| **16-19** | 4 | 9 |
| **20-29** | 23 | 49 |
| **30-44** | 32 | 27 |
| **45-54** | 13 | 9 |
| **55-59** | 6 | 2 |
| **60-64** | 6 | 2 |
| **65-74** | 8 | 1 |
| **75+** | 6 | 2 |
| | **100** | **100** |

The analysis of borrowing at Sandwell and Hounslow found that most heavy borrowers were aged 60 and over[46].

Children use public libraries more than adults, with use peaking at about age 10 to 12, when children become more independent in their reading choices. Reading and library usage then declines during teenage years. People rejoin the library in their twenties, but then use libraries sporadically. Once retired, people are more likely to use libraries regularly, and the over sixties may account for a disproportionate amount of library use.

## Library use and ethnic background

A recent study of public libraries and ethnic minorities concluded that:

> 'A social distance exists between the public library and ethnic minority communities which tends to exclude ethnic minority citizens whilst preserving professional autonomy.'[47]

However, the quantitative data in the same study indicates that ethnic minorities use public libraries as much as white people[48]. *Children's reading choices* found that 'Book borrowing from libraries by speakers of languages other than English is higher than in the sample as a whole'[49]. Data collected from those joining Westminster Libraries suggests that people from ethnic minority backgrounds use libraries in proportion to their numbers in the population (Table 9).

Approximately 25% of Westminster's population are from ethnic minorities and they form 24% of adults joining the library. (The sample of those joining a library probably underestimates the participation of ethnic minorities as it only included adults, but 30% of Westminster's ethnic minorities are under 20 years old.)

Studies from America have indicated that education rather than race, has been the strongest predictor of library use. The *Public Library Inquiry* found that black Americans used libraries less than whites, but they had received less schooling[50]. Between 1949 and 1975:

> 'Four of the studies that tested the relationship found that race was not a statistically significant predictor of library use.'[51]

## What are public libraries used for?

Public libraries are mostly used for borrowing books. The 'circulation of books is the single most-used service for the public library'[52]. The Berkshire survey found that 86% of customers were borrowing, returning or requesting books; in Hillingdon 78% of users were borrowing or returning books[53]. Westminster's customer survey found that 67% of users returned or borrowed books[54]. It is possible, given the wording of the survey, that more people used the library hoping to borrow books but did not do so.

However most studies have found the single greatest source of dissatisfaction with the library service, to be the book stock. The Berkshire study revealed that 'the overall standard of the book stock in particular is causing considerable concern to customers'[55].

The 1998 *Books and the consumer* survey asked about opinions of public libraries, but:

> 'Only two things, range of adult fiction books for loan, and range of adult non fiction books for loan, were answered

*Table 9. Ethnic minority use of Westminster Libraries*

| | **% of population 1996** | **% of those joining library 1997** | **% of users (1996 survey)** |
|---|---|---|---|
| White | 75 | 76 | 74 |
| Black | 8 | 6 | 8 |
| Indian | 2 | 2 | 3 |
| Pakistani | 1 | 3 | 1 |
| Bangladeshi | 2 | 1 | 1 |
| Chinese | 2 | 4 | 4 |
| Other Asian | 3 | 4 | 3 |
| Other | 6 | 5 | 5 |
| **All ethnic minorities** | **25** | **24** | **26** |

by sufficient numbers to allow for sensible analysis.'[56]

A large survey of users in Lincolnshire and the North East concluded 'to most library users, of all ages and groups, the library is primarily a means of recreation'[57]. The study of library use in Sheffield concluded '89% of community library users use the library for leisure purposes'[58].

## Concentration of book use

Analysis of the use of adult book stock in Westminster Libraries from 1 October 1996 to 30 September 1997, shows the number of issues of each item in stock during that period. Taking an average number of issues for each issue range (3.5, 8, 13 and 22 respectively) it is possible to estimate the distribution of issues.

A small number of items account for most of the issues (Table 10).

A similar investigation of book stock use was based on Huddersfield Library in 1980[59]. Although the Huddersfield data does not show the same degree of bias in the use of the book stock, the most popular 27% of items accounted for 60% of issues. Demand is concentrated on a small range of items; many people want to read the same books.

## Fiction borrowing

The *Public Library Inquiry* concluded that most demand was for current, or recently published, fiction. Data from private rental librar-

*Table 10. Distribution of issues and stock – Westminster*

| | 0 issues | 1 issues | 2-5 issues | 6-10 issues | 11-15 issues | 16+ issues |
|---|---|---|---|---|---|---|
| Fiction | 24 | 6 | 27 | 26 | 12 | 5 |
| Non fiction | 31 | 9 | 28 | 20 | 9 | 4 |
| Music Books/Scores | 47 | 11 | 25 | 12 | 4 | 1 |
| **% of items – total** | **30** | **8** | **27** | **21** | **9** | **4** |
| **% of issues – total** | **0** | **2** | **20** | **35** | **26** | **18** |

*Percentage of stock generating indicated number of issues – 1996/97.*

*Table 11. Use of non fiction in Westminster*

| Rank | Subject | % on loan | Issues/item | Popularity index | Median rank – all libraries |
|---|---|---|---|---|---|
| 1 | Computers | 80 | 9.7 | 772 | 1 |
| 2 | Law | 69 | 7.4 | 513 | 5 |
| 3 | Language | 58 | 8.6 | 497 | 5 |
| 4 | Paranormal | 49 | 9.0 | 440 | 5 |
| 5 | Business | 55 | 7.8 | 428 | 6 |
| 6 | Body | 50 | 8.5 | 424 | 5 |
| 7 | DIY | 50 | 7.7 | 381 | 5 |
| 8 | Photography | 46 | 7.2 | 328 | 11 |
| 9 | Crime | 39 | 8.1 | 321 | 10 |
| 10 | Food & Drink | 41 | 7.5 | 309 | 11 |
| 31 | History | 23 | 5.0 | 115 | 28 |
| 32 | Sport | 27 | 4.2 | 112 | 29 |
| 33 | Music | 28 | 3.9 | 110 | 30 |
| 34 | Environment | 22 | 4.4 | 97 | 31 |
| 35 | Nature | 19 | 4.2 | 79 | 33 |
| 36 | Literature | 23 | 2.7 | 61 | 34 |
| | **Non fiction** | **35** | **5.8** | **205** | |

*Popularity index = product of % on loan and issues/item.*

ies showed that 77% of circulation had been published in the previous five months, and 90% of circulation was published in the previous year. The records of Montclair Library, New Jersey, showed that half the fiction circulation had been published in the previous year[6]. Analysis of fiction borrowing in Westminster supports this view. Of the total fiction stock, 35% is on loan at any one time. Of 30 hardback and paperback best-sellers of 1997, which were in stock, between 80% and 90% of all copies were on loan. Even where 50 or 60 copies of a title were in stock (*The Beach*, *High Fidelity*), over 80% were on loan. *Bridget Jones's Diary* was top of the best-seller list from June 1997 through into 1998 and sold over one million copies, but even when Westminster Libraries increased their holdings to 90 copies, almost all were always on loan. There is heavy demand for a small number of titles, a conclusion supported by Goodall's review of browsing in libraries[61].

There is also little difference between what is bought and what is borrowed; people use the most convenient means of acquiring the books they want to read.

## Non-fiction borrowing

Westminster Libraries' non-fiction stock is divided into 36 reader interest categories, and the intensity of use of these categories was compared in some detail.

Table 11 shows the results of a comparative analysis of the use of non-fiction at all Westminster Libraries. Combining the percentage on loan with the average issues per item gives a single value that allows all subjects to be ranked. All measures of use show a remarkable consistency across all libraries. Whatever measure is used, the same type of books are popular at all libraries – principally books on computers.

The most heavily-used categories of non-fiction are practical subjects, rather than academic material. *Language* is foreign language learning, mostly basic cassette courses; *Law* and *Business* contain books on writing wills, renting and letting property, legal textbooks, management, accounting, CV's and interview skills, marketing and workplace issues. *Body* is health and psychology, including self-help. Books are wanted as 'tools'.

These results confirm Sharon Baker's supposition, that practical books are used much more extensively than academic books[62]. As 80% of *computer* books and almost 70% of *law* books are on loan, there is probably considerable unmet demand.

*Table 12. Westminster's most borrowed non fiction titles by subject*

| | | Copies issued 16+ times | | |
|---|---|---|---|---|
| | **% of total non fiction stock** | **Number** | **As % of most borrowed stock** | **As % of subject stock** |
| Computers | 2.6 | 1,122 | 15.0 | 39.1 |
| Body | 4.9 | 1,495 | 20.0 | 28.0 |
| Law | 1.2 | 215 | 2.9 | 16.4 |
| Language | 2.6 | 445 | 6.0 | 15.7 |
| DIY | 1.3 | 201 | 2.7 | 13.9 |
| Business | 5.9 | 680 | 9.1 | 10.5 |
| Paranormal | 1.3 | 144 | 1.9 | 10.5 |
| History | 9.4 | 167 | 2.2 | 1.6 |
| Society | 4.2 | 68 | 0.9 | 1.5 |
| Literature | 10.7 | 160 | 2.1 | 1.4 |
| Parenting | 1.3 | 18 | 0.2 | 1.3 |
| Myths | 0.5 | 3 | 0.0 | 0.5 |
| Environment | 0.5 | 0 | 0.0 | 0.0 |
| **Non fiction** | **100.0** | **7,475** | **100.0** | **6.9** |

A list was extracted from the circulation database, consisting of all non fiction titles which had been borrowed at least 16 times in the year. These approximately 7,500 items (5% of the total) accounted for about 20% of the issues. Table 12 shows the top and bottom of the list analysed by subject.

The same practical subjects already shown to be most in demand, also provide the most borrowed titles. The individual titles in the list are almost all practical books required for immediate concerns – life skills, writing CV's, personal computers, grammar, travel guides, the driving test. Of the *Literature* titles in the list almost all are books on practical writing skills and creative writing; the remainder are school texts, such as Shakespeare. Other titles to feature are recent best-selling biographies, classic scientific, literary and cultural texts and other recent best-sellers – Nick Hornby, Bill Bryson.

Summarising a survey of issues in 1960, the librarian of Folkestone wrote:

> 'there is an ever increasing call for books as tools: for that diversified class of simple, unpretentious, informative manuals, which help readers to play bowls, arrange flowers, decorate the house, keep tropical fish, and the like, more effectively'[63].

Little has changed in 40 years. Academic subjects provide few of the titles in the most borrowed list. *Art*, *Literature*, *History*, *Society*, and *Nature* are 30% of the book stock but only 7% of the most borrowed titles.

Most Westminster Library members only borrow books once or twice a year, so probably do so only when they have a particular need. The study of use at Sandwell and Hounslow found that heavy borrowers were more likely to borrow mostly fiction; light borrowers were more likely to borrow non-fiction[64]. The pattern of use looks like a minority of users regularly borrowing a lot of fiction, but the majority of users casually borrowing one or two non-fiction items as and when needed.

## Music and video borrowing

A study of the use of audio-visual materials in Westminster found that people borrowed much the same videos from the public library as they did from commercial outlets[65].

Most videos borrowed from the library are current adult feature films. Educational videos are a small percentage of the total video issue. Most children's video issues are of cartoons and Disney films.

*Table 13. Westminster video issues*

| Type of video | % of video issues |
|---|---|
| Children's | 15 |
| Feature films | 71 |
| Other (fitness, travel, sport etc.) | 14 |

*1996/97*

Heavily promoted new release feature films, such as *Independence Day*, were also profitable for the public library service. People do not come to the public library for alternative material to the high street commercial outlets. If the library is convenient for what they want – current adult feature films to view at the weekend – they will use that service. People borrow the same music CDs and cassettes from public libraries as are sold in shops. Of all albums sold, 56% are rock and pop music; 54% of Westminster's music issues were rock and pop. Top 40 albums are about one quarter of all album sales. The same material is in demand from the public library. In July 1997 an analysis of all copies of top 40 UK albums in the library's stock, found that 85% were on loan. Of the 1996 top 40 albums, two thirds of all copies were on loan. On average one third of the library's music stock is on loan[66].

## Customer requests

Details of all requests made at Westminster Libraries over a two-week period in 1999 were collected, giving a sample of 1,931 requests. 80% of items requested were already in the library's stock. Half the requests were for adult non fiction books, which were categorised by type of book, as either recent publications, practical books or academic books.

*Table 14. Type of non fiction requested – Westminster*

| Type | % |
|---|---|
| New / best-seller | 30% |
| Practical / textbook | 37% |
| Academic | 32% |

About 90% of non fiction requests are for recently published items, items already in stock, or items for which there is likely to be reasonably extensive demand (e.g. travel guides, health, computers).

Titles published within the last year were 57% of fiction requests. Most requests for items published more than two years ago, were for established classics (*Lord of the Rings*, *Catcher in the Rye*), books of recent films (*Beloved*), or older works by popular authors (David Lodge, Jilly Cooper). Most fiction requests are for mainstream and predictable items. As with adult fiction borrowing in general, there is heavy demand for new and recently published titles. Almost all music requests were for current rock and pop CDs. Only three requests were for classical music not already in stock.

In 1997/98 requests for journal articles were 0.2% of all requests, but three quarters of these were from the same person. About 20 people a year use Westminster Libraries to order journal articles. The library requests service is used almost entirely to order mainstream material – new publications, current CDs, popular non fiction. Few people use the service for research purposes or to gain access to wider library holdings, for example at the British Library.

## Conclusion

Significant aspects of public library use have remained constant for much of their history:

- their use for leisure and recreational reading;
- their use by a minority of the population;
- their greater use by more educated people;
- the close relationship between reading, book ownership and library use;
- that women use libraries more than men.

Library membership records suggest that respondents to surveys over-estimate their own library use, and that only one third of the population belong to a library, with the rate likely to be higher in more affluent areas. Most people do not maintain continuous library membership, but join for a period when it is convenient or useful. Two thirds of the population have this casual relationship to the library service. Most library members do not use a library regularly, but do so casually, when a need arises, perhaps only once or twice a year. About 10% of the population use libraries at least once a month, and they account for most library borrowing.

The amount of education is the single greatest determinant of library use. The more educated a person, the more likely they are to use libraries. It is therefore unlikely that public libraries affect the educational level of the population, rather they reinforce the existing situation. The amount of education is closely related to social class and wealth, and low social class and economic status invariably indicate less use of libraries. Increasing the average terminal education age is the best way to increase public library use.

Women use libraries more, and borrow more books, than men. School age children use libraries more than adults. Most adults either use libraries regularly or very occasionally, but children are more likely to be intermediate users – four to six times a year. There is some evidence that adult regular users are more likely to be 60 and over. People from ethnic minorities use libraries as much as white people.

Most library use involves the borrowing of books; users' most significant concerns relate to the quality of the book stock and the availability of titles in demand. Most library use is for leisure reading or for help solving immediate personal or life issues (e.g. travel guides, using a personal computer, writing a CV, self-help manuals) – 'books as tools'. Most demand is for current fiction or for up-to-date books of immediate practical use. There is little evidence of public library use for research or independent education.

Demand is concentrated on a small number of subjects and titles. There is considerable similarity in the demands made of public libraries and commercial book, music and video stores. Libraries are not perceived as offering a different product to bookshops, but they may be more or less convenient to use. Library users are more likely to buy books than non-users, and book buyers are more likely to be library users.

Educated people who have most access to information, already use public libraries more than those who are less educated. It is not yet clear that information technology and the Internet are changing the nature of public library use. Past evidence suggests that new services mostly attract existing users.

The wealth of quantitative data available from library systems provides a useful corrective to over reliance on consumer surveys, and can give a more accurate picture of library use. Given the range, quality and consistency of demographic studies, libraries should consider carefully before re-inventing the wheel with further customer surveys. Westminster has a fairly representative clientèle from Inner London and the home counties, but further studies of borrowing patterns in other authorities would be useful to establish the extent of similar patterns of use.

## References

1. Raber, D. 'Professional ideology and conflict: reactions to the Public Library Inquiry.' In: Greiner, J. (ed.) *Research issues in public librarianship: trends for the future.* (Westport, Conn.: Greenwood Press, 1994, pp.31-50.)

2. Berelson, B. *The library's public.* (New York: Columbia University Press, 1949.)

3. Groombridge, B. *The Londoner and his library.* (London: Research Institute for Consumer Affairs, 1964.)

4. Luckham, B. *The library in society: a study of the library in an urban setting.* (London: Library Association, 1971, p.xii.)

5. Totterdell, B. and Bird, J. *The effective library: report of the Hillingdon project on public library effectiveness.* (London: Library Association, 1976.)

6. Ward, M. *Readers and library users.* (London: Library Association, 1977.)

7. *Borrowed time? The future of public libraries in the UK.* (Stroud: Comedia, 1993, p.9.)

8. *Review of the public library service in England and Wales for the Department of National Heritage: final report.* (London: Association for Information Management, 1995.)

9. *Household library use survey 1998.* (London: Book Marketing Ltd., 1998.)

10. Kenyon, F. *Report on public libraries in England and Wales.* (London: HMSO, 1927, p.69.)

11. McColvin, L. *The public library system of Great Britain: a report on its present condition with proposals for post-war reorganisation.* (London: Library Association, 1942, p.10.)

12. Library Association. *Press releases. http://www.la-hq.org.uk/directory/press_desk/199907.html* (visited 25 February 1999)

13. Chartered Institute of Public Finance and Accountancy. *Public library statistics 1997-98 actuals.* (London: Institute of Public Finance, 1998.)

14. *Books and the consumer 1994-1996: a comprehensive guide to the library market.* (London: Book Marketing Ltd., 1997, p.21.)

15. Westminster Libraries Business Unit. *Community profile.* (London: Westminster Libraries, 1997, p.36.)

16. Luckham, *The library in society: a study of the library in an urban setting.* (London: Library Association, 1971, pp.21-23.)

17. Totterdell, B. and Bird, J. *The effective library: report of the Hillingdon project on public library effectiveness.* (London: Library Association, 1976, p.57.)

18. Proctor, R., Usherwood, B. and Sobczyk, G. *What do people do when their public library service closes down? An investigation into the impact of the Sheffield Libraries strike.* British Library Research and Development Report 6224 (Boston Spa: British Library, 1996, p.5.)

19. Luckham, B. *The library in society: a study of the library in an urban setting.* (London: Library Association, 1971, p.50.)

20. Groombridge, B. *The Londoner and his library.* (London: Research Institute for Consumer Affairs, 1964, p.21.)

21. Thorpe, S. 'Market research in Enfield Libraries.' *Public Library Journal* 1(4), 1986, pp.49-54.

22. Westminster Libraries Business Unit. *Libraries user survey – full report.* (London: Westminster Libraries, 1996, p.26.)

23. Berelson, B. *The library's public.* (New York: Columbia University Press, 1949, p.101.)

24. Westminster Libraries Business Unit. *Libraries user survey – full report.* (London: Westminster Libraries, 1996, pp.12, 26.)

25. England, L. *Borrowing books: readership and library usage in Great Britain.* (London: Book Marketing Ltd., 1992, p.18.)

26. *Libraries and the consumer 1994-1996, a comprehensive guide to the library market.* (London: Book Marketing Ltd., 1997, p.32.)

27. England, L. and Sumsion, J. *Perspectives of public library use: a compendium of survey information.* (Loughborough: Library and Information Statistics Unit and London: Book Marketing Ltd., 1995, pp.123-132.)

28. Berelson, B. *The library's public.* (New York: Columbia University Press, 1949, p. 16.)

29. Groombridge, *The Londoner and his library.* (London: Research Institute for Consumer Affairs, 1964, p. 44.)

30. Luckham, B. *The library in society: a study of the library in an urban setting.* (London: Library Association, 1971, p. 43.)

31. Zweizig, D. and Dervin, B. 'Public library use, users, uses.' *Advances in librarianship* 7, 1977, p.241; Granheim, E. 'The library user survey for 1988. Does it tell us anything new?' *Scandinavian Public Library Quarterly* 23(2), 1990, p.11; England, L. *Libraries and the consumer 1994-1996: a comprehensive guide to the library market.* (London: Book Marketing Ltd., 1997, p.31); Yilmaz, B. 'A sociological study of public library use in Ankara, Turkey.' *Journal of Librarianship and Information Science* 30(4), 1998, p.266; Hall, C. and Coles, M. *Children's reading choices.* (London: Routledge, 1999, p.120.)

32. Jackson, A. and Martin, P. 'The non-use of Manchester's library service: an investigation.' *Public Library Journal* 6(4), 1991, p.113.

33. Sumsion, J. *PLR in practice: a report to the advisory committee*. (2nd edition) (Stockton-on-Tees: Registrar of Public Lending Right, 1991, p.139.)

34. *Reading in Tottenham: a report on a survey carried out by Mass Observation on behalf of the Tottenham Borough Council.* (London: Tottenham Public Libraries and Museums, 1952, p.14.)

35. Berelson, B. *The library's public.* (New York: Columbia University Press, 1949, p.28-30.)

36. Groombridge, B. *The Londoner and his library.* (London: Research Institute for Consumer Affairs, 1964, p.17.); Luckham, B. *The library in society: a study of the library in an urban setting.* (London: Library Association, 1971, p.26.); Totterdell, B. and Bird, J. *The effective library: report of the Hillingdon project on public library effectiveness*. (London: Library Association, 1976, p.49.); *Review of the public library service in England and Wales for the Department of National Heritage: final report.* (London: Association for Information Management, 1995, p.118.); England, L. *Libraries and the consumer 1994-1996: a comprehensive guide to the library market.* (London: Book Marketing Ltd., 1997, p.15.); Yilmaz, 'A sociological study of public library use in Ankara, Turkey.' *Journal of Librarianship and Information Science* 30(4), 1998, p.264.

37. Zweizig, D. and Dervin, B. 'Public library use, users, uses.' *Advances in Librarianship* 7, 1977, p.232; Lange, J. 'Public library users, non-users, and type of library use.' *Public Library Quarterly* 8(1/2), 1987/88, p.51.

38. Taylor, J. and Johnson, I. *Public libraries and their use: a research report on the use of public library buildings with implications for their distribution, location and design.* (London: HMSO, 1973, p.9.)

39. Stevens, A. *Getting to know our customers: report of the findings of a county-wide user survey, 25 September – 7 October 1989.* (Reading: Library and Information Service, Royal County of Berkshire, 1990, p.11.)

40. England, L. *The library user: the reading habits and attitudes of public library users in Great Britain.* British National Bibliography Research Fund Report 68. (London: Book Marketing Ltd., 1994, p.9.)

41. Hall, C. and Coles, M. *Children's reading choices.* (London: Routledge, 1999, pp.2, 119.)

42. England, L. *Libraries and the consumer 1994-1996: a comprehensive guide to the library market.* (London: Book Marketing Ltd., 1997, p.31.); England, L. and Sumsion, J. *Perspectives of public library use: a compendium of survey information.* (Loughborough: Library and Information Statistics Unit and London: Book Marketing Ltd., 1995, p. 127.)

43. Berelson, B. *The library's public.* (New York: Columbia University Press, 1949, pp.19-24.)

44. Hall, C. and Coles, M. *Children's reading choices.* (London: Routledge, 1999, pp.2, 118.)

45. *Household library use survey 1998.* (London: Book Marketing Ltd., 1998, p.18.)

46. England, L. and Sumsion, J. *Perspectives of public library use: a compendium of survey information.* (Loughborough: Library and Information Statistics Unit and London: Book Marketing Ltd., 1995, p.127.)

47. Roach, P. and Morrison, M. *Public libraries, ethnic diversity and citizenship* (London: British Library, 1998, p.7.)

48. Roach, P. and Morrison, M. *Public libraries, ethnic diversity and citizenship* (London: British Library, 1998, pp.55, 71, 141.)

49. Hall, C. and Coles, M. *Children's reading choices.* (London: Routledge, 1999, p.107.)

50. Berelson, B. *The library's public.* (New York: Columbia University Press, 1949, p.40.)

51. Zweizig, D. and Dervin, B. 'Public library use, users, uses.' *Advances in Librarianship* 7, 1977, p.240.

52. Berelson, B. *The library's public*. (New York: Columbia University Press, 1949, p.52.)

53. Stevens, A. *Getting to know our customers: report of the findings of a county-wide user survey, 25 September – 7 October 1989.* (Reading: Library and Information Service, Royal County of Berkshire, 1990, p.29.); Totterdell, B. and Bird, J. *The effective library: report of the Hillingdon project on public library effectiveness.* (London: Library Association, 1976, p.47.)

54. Westminster Libraries Business Unit. *Libraries user survey – full report.* (London: Westminster Libraries, 1996, p.7.)

55. Berelson, B. *The library's public*. (New York: Columbia University Press, 1949, p. 83.); Totterdell, B. and Bird, J. *The effective library: report of the Hillingdon project on public library effectiveness.* (London: Library Association, 1976, p.113.); Stevens, A. *Getting to know our customers: report of the findings of a county-wide user survey, 25 September – 7 October 1989.* (Reading: Library and Information Service, Royal County of Berkshire, 1990, p.34.)

56. *Household library use survey 1998.* (London: Book Marketing Ltd., 1998, p.101.)

57. Taylor, J. and Johnson, I. *Public libraries and their use: a research report on the use of public library buildings with implications for their distribution, location and design* (London: HMSO, 1973, p.62.)

58. Proctor, R., Usherwood, B. and Sobczyk, G. *What do people do when their public library service closes down? An investigation into the impact of the Sheffield Libraries strike.* British Library Research and Development Report 6224 (Boston Spa: British Library, 1996, p.22.)

59. Bagust, A. 'A circulation model for busy public libraries.' *Journal of Documentation* 39(1), 1983, pp.24-37.

60. Berelson, B. *The library's public*. (New York: Columbia University Press, 1949, p.62.)

61. Goodall, D. *Browsing in public libraries.* (Loughborough: Library and Information Statistics Unit, 1989, p.65.)

62. Baker, S. *The responsive public library collection: how to develop and market it* (Englewood, Colo.: Libraries Unlimited, 1993, p.55.)

63. Kelly, T. *A history of public libraries in Great Britain 1845-1975*. (2nd edition) (London: Library Association, 1977, p.384.)

64. England, L. and Sumsion, J. *Perspectives of public library use: a compendium of survey information.* (Loughborough: Library and Information Statistics Unit and London: Book Marketing Ltd., 1995, p.128.)

65. Westminster Libraries Business Unit. *Audiovisual materials.* (London: Westminster Libraries, 1997, pp.26-33.)

66. Westminster Libraries Business Unit. *Audiovisual materials.* (London: Westminster Libraries, 1997, pp. 20-23, 80-81.)

# The net generation: the experiences, attitudes and behaviour of children using the Internet for their own purposes

*Peter Williams*

*Internet Studies Research Group, Department of Information Science,*
*City University, Northampton Square, London EC1V OHB*
*Email: pw@is.city.ac.uk*

**Abstract** *The much trumpeted 'information society' borne by the Internet has been greeted by a confusing mixture of delight and trepidation, not least regarding the attendant implications for children. No clear consensus has emerged among librarians, educators or others as to whether the potential benefits of the system outweigh any risk associated with retrieving potentially harmful material. Presented here is a pilot research study that examines how children use the Internet, and what their opinions of the system are. This was undertaken to inform a research proposal to be submitted to funding organisations. Results of the preliminary fieldwork point to much independent educational use, particularly with regard to project work. Printouts of photographs and documents were prized more than the intangible information itself, so children were not discouraged by the reading level of a text and did not evaluate relevance on this criterion. Leisure use was dominated by game playing. Some children accessed online chat groups, with occasional foul language encountered accepted as a natural, if unwanted, consequence of Internet use. There was little evidence of other potentially harmful or upsetting material being viewed.*

## Introduction

Despite many initiatives promoting the Internet as a potentially excellent tool for teaching and learning, including those by the Government[1] and the BBC and Library Association jointly[2], the phenomenon has not been greeted with universal acclaim or enthusiasm regarding its use by children – a major source of disquiet being the accessibility of material perceived as harmful. Many recent news stories[3,4,5] have only reinforced these fears.

With such tremendous dichotomy between the Internet's potential to enrich or deprave lives, and with some individuals themselves[6] ambivalent over the prospects, the time is surely right for an in-depth investigation that finally provides hard information on Internet usage, experience and attitudes of children themselves. In response to this challenge, the Internet Studies Research Group at City University are currently exploring lines of future research. This will examine and analyse the way children *really* exploit the Internet, what they think of it, the extent of their leisure time use, and the benefits, enjoyment or despair they consider it to have given them. The emphasis will not concentrate purely on any negative aspects of Internet use, although experiences and reaction of users to content that may be considered harmful either by themselves or supervising adults will be examined within the broad framework of general use patterns. This paper discusses the results of some preliminary fieldwork.

## Background and literature review

Although much work has been undertaken on how children use information technology, the research and commentary so far specifically regarding the Internet has focused on various educative possibilities[7,8] teaching children to be discerning users[9], controlled information retrieval performances of children[10], or issues such as gender differences[11]. Apart from the rather market led research of the public opinion organisation NOP[12], which plans to collaborate with the proposed study, children's spontaneous use of the Internet and their views and attitudes towards it hardly appear to have been addressed. Indeed, Kafai *et al*[13] provide possibly the only study, apart from the NOP, which actually seeks children's own opinions. Even here, however, the context was highly controlled, and included researchers accessing pre-determined sites for students to evaluate and index. Children placed much emphasis on 'visual relevancy' and had little interest in reading textual information. The researchers suggest this shows that many sites found by free searching were inappropriate for the age group, and that children's growing familiarity with multimedia gave them a predisposition for audio and visual material regardless of the suitability of written information. Children appeared eager to use the Internet for school work (numbers who actually did so were not given) and were, not surprisingly, keen to have their own work published online.

In the NOP study, 'Kids.net', researchers asked child respondents several set questions about their Internet use. Results indicated that almost half (46%) of children have used the Internet, with research for school projects and educational material 'one of the most popular applications'. Seventy-nine per cent of children consider that the Internet helps with their learning. Away from school work, a major activity was browsing the Web for products to buy – no less than 17% have actually made online purchases, generally by using their parents' credit cards. One interesting finding was the popularity of 'multitasking'. Fewer than one in five young users do nothing else whilst online. Other activities undertaken simultaneously include listening to music (59%), watching television (23%) and reading magazines (13%)[14].

The issue of unsolicited material forms a major part of the ongoing study. More than half of 4,000 respondents in a questionnaire survey have been 'taken aback' by something they have accessed on the Internet, no less than 20% been embarrassed or upset, and nearly 3% found something that actually frightened them[15]. Tabloid journalists were only too keen to publish these findings, with *The Mirror's* '500,000 children upset by the Internet'[16] typical. Interestingly, much of the material was not of the pornography/racism type commonly thought of with regard to possible negative effects (although this aspect was not reported). One nine year old, for example, reported his distress as an animal lover at seeing pictures of otter culling[17].

These findings only highlight the need for an in-depth investigation that studies children's use of the Internet, their opinions about it, and the incidence of their accessing unwelcome material.

## Aims and objectives

The general aims of the pilot project were to:

- explore the activities children engage in on the Internet, their search behaviour, what they want from the Internet and how it fits into their lives; and
- investigate the problem of potentially harmful or upsetting material on the Internet that may prejudice its full exploitation.

More specifically the project sought to:

- obtain a feel for the number of users and extent of use (precise quantative data was not possible given the small sample size and specific geographic location);
- examine any differences between home and school use and discriminate between study/leisure activities;
- explore sites visited, their attractions, faults and relevance (in the eyes of the children);
- examine searching/browsing behaviour and whether this varies according to task;
- compile instances of the access of material children or their carers consider to be undesirable, upsetting or harmful;
- assess the nature, extent and depth of disquiet these experiences have given rise to;
- compare findings regarding all of the above with the opinions and attitudes of teachers and possible use policies adopted for children.

## Context, sample and methodology

To aid the preparation of fieldwork, informal visits were undertaken at a primary school situated in the Bengeo area of Hertford. Although fairly affluent in terms of home/car ownership etc. (above average for the county and well above the national average, according to census data[18]) staff refuted suggestions that the school was 'middle class', pointing to the wide catchment area which included various estates that could not be labelled thus. The researcher spent one day a week, for a period of six weeks, with a class of 10 and 11 year olds (i.e. top juniors) in the ostensible capacity of classroom assistant. In addition to undertaking interviews and observing educational use of the Internet, help was given to the class teacher in the form of small group teaching and general supervision.

There were many advantages in becoming immersed in the environment of the school rather than simply concentrating on formal data gathering. Firstly, being engaged in general classroom activities (particularly as an ex-teacher) helped gain the confidence of both pupils and teachers, develop a rapport and demonstrate an understanding of the teaching/learning process. Constant interaction also aided the projection of the researcher as a dispassionate and neutral figure whose interest was to explore and discover rather than sit in judgement. It also afforded opportunities to observe the use generally of IT in the classroom and in the lives of the children, and thus contextualise exploitation of the Internet. Finally, the extensive presence in the school permitted data collection by many informal and ad hoc methods beyond the 'formal' interviews undertaken with small groups of two or three children. These included:

- informal 'chats' with pupils at break and lunch times;
- in passing during the course of other activities (for example, printed web pages were often proffered as examples of Internet research);
- from members of staff, informally, in the staff room.

These wide ranging and opportunistic data gathering practices although extremely effective in accruing information, made spontaneous recording of comments etc. virtually impossible and it was also difficult to keep track of numbers spoken to. This was not considered to be a major problem, however, as the study did not set out to be a quantitative analysis. Research findings are indicative. Thus if only one respondent cited a particular use/experience, this is considered important. Nevertheless, at the end of the sixth fieldwork session a straw poll was taken of those who (a) had the Internet at home and (b) had experience of it. Only two of the latter had not contributed to the fieldwork, and they were immediately interviewed. Other children in other classes also contributed to the fieldwork, such as those met at break/lunch times, so it is impossible to enumerate respondents. Notes were taken during the course of more formal interviews, undertaken at a specific location in one corner of the classroom (an open artwork area was made available for this purpose) and as soon as possible after any informal interaction.

## Preliminary fieldwork: results

### *Users*

The figure of 37 children (out of classes totalling 60) already having experience of the Internet reflects the growing intrusion of the technology into their lives. Indeed, even non-users appear to have a fair notion of what the Internet is and the different ways it can be exploited. Many claimed to be on the verge of connection at home and it was easy to discern an element of peer pressure here. Of course, the extent of playground talk about the Internet, CD-Roms and computers in general contributing to any possible pressure to 'go online', occurring in the absence of adults is not known. Within the small sample surveyed there were no gender differences with regard to user numbers.

### *Use*

No clear pattern emerged regarding time online or specific uses of the Internet, either in terms of school or leisure pursuits. Children's estimates ranged from 1% to 99% of use being for leisure, and vice versa for education, although the latter did appear to be more dominant and was generally the first mentioned. Of course, discussing use within the physical environment of an educational establishment and, moreover, to an adult – even one not formally labelled as a teacher – may have brought these experiences to mind first. Nevertheless, the accounts given and printed pages offered do point to considerable pedagogical exploitation, supporting Kafai *et al's*[13] finding that children would be happy to use the Internet for study purposes.

School use included searching for images to down-load, finding or checking facts, and printing off text material. The current year six project on 'endangered species' lent itself, it was generally agreed, to such Internet research. Interestingly, although much general class work was undertaken by children in word-processed form, no one mentioned copying and pasting text from web sites straight onto a word-processing application. When this possibility was suggested to five or six children over school lunch, the idea was dismissed. 'They'd know it wasn't yours!' ... 'It would be too long' ... 'We don't need to.'

Someone suggested that a printout itself, if included with students' own work, would be acceptable as part of a project in much the same way that cuttings from magazines, pictures and photocopies might be. Thus the information itself was for many children secondary to the physical medium. Indeed, there was much evidence that printouts were not regarded as being for comprehensive reading – capturing such a document was an end in itself and not a source from which information had to be extracted. Many children made comments such as 'you only read the facts' ... 'I just read bits' ... 'I take what I need'. Perhaps for this reason, when pressed on the subject of suitability of material in terms of reading level, comprehension etc. there was universal agreement that required material, even if not specifically written for a young audience, abounded on the Internet and was relatively easy to find. An examination of how children extract key facts from lengthy web pages may form an interesting part of a the full research project.

Of course, study resources on the Internet do not consist purely of passive text documents or photographs. The BBC education site, to give one example, includes various number games for children. Typically imputing correct answers to basic arithmetic questions permits the 'player' to undertake online operations such as negotiate a maze or score basketball points etc. This kind of activity is clearly educational. The knowledge or skills acquisition, however narrowly defined that acquisition might be, can be measured by achieving time targets or obtaining higher scores. The children's lack of familiarity with such sites may speculatively be indicative of a number of factors in (at least current) Internet use. These can be summarised as:

- lack of awareness of the facilities available;
- casual use where general activity sometimes leads to useful information being around, rather than the positive seeking out of educational sites and activities containing such educational games;
- a reluctance to engage in such overtly educational activities – it may be more fun to surf; and
- the redundance of the Internet in the face of CD-Rom based educational games activities.

Leisure use did include extensive game playing, however, and (predictably for the teachers, but less so for the non-streetwise interviewer) learning solutions to games from 'game cheating' sites. Exploiting the Internet to download free 'demos' of games was also common practice. With the usual caveat that the sample was rather small, there was a discernible attitude difference here between boys and girls, with the former being far more enthusiastic about this facility. No one described undertaking any such activity with a remote partner, although communication in other contexts was practised. Indeed, about 50% of Internet users have sent email, and so called 'chat rooms' were also used by a minority. These were of the bulletin board type. No one mentioned the 3D 'rooms' populated by 'Avatars' that it is possible to delve into. Use was ill-defined, but indications are that contacts were fleeting and superficial. Message boards were seen as areas to drop into out of curiosity, to see what was being said, and leave again. No attempts appear to have been made to forge longer term relationships. Many of those who had not used the Internet for this type of communication were as unaware as non-users as to the setting up and practice of 'chatting'.

Email was exclusively used by those who already had friends or – more commonly –

family overseas. As in the case of message boards, 'pen friends' had not been acquired through the Internet. One heavy user of email had lived in the USA for two years and kept in contact with American friends while other regular users had relatives in Spain and Canada respectively.

Although use of the Internet is facilitated at school, access is limited as yet due to the system only just having been installed, and largely on an ad hoc basis. Perhaps not surprisingly, therefore, up to the present time every one of those surveyed had been introduced to the system at home. This was generally by parents who needed computers for work related activities. A small minority claimed the equipment had been purchased exclusively or principally for their and/or other siblings' use. Initial personal Internet experience and training/instruction were also undertaken at home. In line with current surveys that suggest increasing use of information technology by women[19,20] it was more often the mother who provided the demonstrations, help etc. and who used a PC for her own activities.

### *Search behaviour*

Very little observation of free Internet searching was undertaken during the pilot study. In interviews addressing this issue there was universal agreement that using the Internet was 'easy', and in response to the question 'How do you find things on the Internet?' the general reply was that 'I just tell it what I want to know'. When pressed one child said he used the same words 'as if I was looking it up in a dictionary'. Another said 'you use the words like in Encarta'. Interestingly some children did not describe imputing search terms, but referred simply to a list of items to choose from on the screen. Explanations from this group as to how they obtained the list were vague and it is not clear whether they were describing a search hit list, a directory list or a bookmark file. A minority said that adults undertook preliminary searches. It is also possible that some parents use formal portal sites or self created 'hot lists' as home pages. These would have the dual advantage of guiding their children to interesting and relevant sites and minimise the risk of accessing inappropriate material.

In describing this retrieval of material many pupils referred to their current project on endangered species. Key words used in searches were, without exception, names of animals etc. that the children knew to be relevant to their projects. No one included either the words 'endangered' or 'species'. Whilst this paper is not concerned at a technical level with search strategies, it is worthwhile speculating that the use of such concrete terms, rather than generics or abstract expressions may actually be the most appropriate way to search the Internet. This is because the vast majority of pages are not indexed by humans and, in many cases, documents may be lost through the practice of using generic terms or concepts rather than the natural language of the text.

Nearly all children were happy with their search results. Only one expressed irritation that so much retrieval material was inappropriate. Others said it was 'fun' to look at 'other stuff' and to follow links through, even if this meant wandering off the subject. The bookmark facility was mentioned, but this appears to have been used more for leisure sites – particularly those offering games. There was no evidence of children building up a library of relevant educational sites to which they turned with further information queries although, as alluded to earlier, hidden parental involvement may have created the 'lists' mentioned in interviews.

### *Accessing inappropriate material*

The issue of harmful or inappropriate material on the Internet, forming an important, if not a major part, of the project proposal was explored. Unlike findings by NOP described earlier, there was little evidence of material being accessed that either had adverse effects on children or that would concern parents or carers. Where pupils were asked if they had accessed material they 'didn't like' or 'that isn't really for you' the common response was to cite pages that did not deal with the topic they were investigating. Much retrieved material was simply 'boring'. The question was not made more specific, i.e. to enquire about violent or abusive language, as the aim of the next stage of the project will be to elicit from children themselves ideas and opinions about material. The fact that only one interviewee even understood the implications of the question without additional prompting suggests that, for this group, any inadvertent viewing of sex, racist or other such sites was inconsequential.

Three specific cases were mentioned where Internet use might have concerned adults, although even here the situations were dismissed, at least by the children, almost as a slight irritation. These include similar experiences reported by two pupils who regularly use the 'chat' facility. Both used groups described as being for their age band, although they were at the younger end, choosing the 10-13 bracket. One described the presence of 'bad language', and the other 'disgusting conversations'. In deference to the children, and cognisant of others who might have been within earshot, no attempt was made to elicit details. Clearly, before the next stage begins a framework needs to be established that provides guidance for researchers in the treatment of young interviewees.

The third example of where potentially unsuitable material was accessed highlights the difficulty for adults both in determining what constitutes harmful material and, consequently, in filtering it out. The offending pages in this case belonged to a US religious cult, and were retrieved from a search on 'iguanas' (made as part of the endangered species project). Apparently – and this caused great amusement on the part of the pupil who found the sites, the cult uses the name iguana to describe itself. The pages were it seems, quite innocuous. The incidence nevertheless shows how more insidious material – possibly more effectual amongst older children – might be innocently retrieved.

### *Teachers' views*

This overall picture of use appeared to be entirely predictable to the teachers. No doubt the constant interaction between pupils and teacher – particularly in the primary school environment, where most of the day is spent with the same class – played a major part in this. Manifestations of Internet use in the form of printouts and photographs etc. also indicated to teachers what children had been researching.

Generally teachers appeared to be positive about the potential of the Internet and felt that those children who had used it had benefited from the experience. Two teachers spoke of the BBC's 'Scavenger Hunt' initiative in which questions are posed whose answers require children to search the Internet. They expressed admiration for how this had been undertaken by the children. Interestingly use of the 'Ask Jeeves' search engine was encouraged, because it allows questions in natural language sentence form, and indexes sites (thus facilitating more generic search terms). One interesting, if peripheral, area for the full project to pursue may be the extent of teaching of search strategies. As teachers are not trained information professionals this may be an area which they either find difficult or, particularly with inexperienced users, are unaware of the possibilities or implications.

Concerns expressed by teachers with regard to their pupils' use of the Internet centred on their own abilities to incorporate the system into their professional work in a way that maximised its potential. Worries about any adverse consequences of children's unsupervised use were secondary to this factor. Typifying the general feeling one teacher opined that there was a general unease that 'the technology is overtaking everyone' and that formal structured training was required. Similarly, a peripatetic special needs teacher was worried that he might not be able to aid children when work he had to help with involved the use of computer facilities. Other teachers also commented that they and/or most of their colleagues were ill-equipped to exploit the technology. Their concern was compounded by a feeling that the Internet had great potential for classroom use, and that therefore expertise was imperative.

The problem of potential access to racist/pornographic or violent material was addressed, but often with regard to older pupils – specifically teachers' own children. One teacher who was enthusiastic about the Internet described what she termed 'the Hitler problem'. This is the well known difficulty of accessing interesting and relevant information on sensitive topics without also retrieving material of, to put it euphemistically, little educational value. She was dubious about filtering software, electing instead to adopt a vigilant approach implying constant adult supervision.

Finally, many points were raised regarding the 'effectiveness' educationally of the Internet. These included whether 'surfing the Net' was valuable in itself, how Internet research differed from book consultation in accessing required information and images, and if it was beneficial to conduct lessons using online materials?

## Conclusion

Both the library research and the small scale (somewhat informal) fieldwork have raised a number of important issues that need to be addressed. For example, in addition to an exploration of sites visited, the question of exactly how each site is exploited needs to be considered. This should include an examination of the information retrieved and the relationship between this and the physical pages downloaded. The NOP finding that the Internet is, indeed, used very extensively for school work provides a good quantative basis, but the qualitative study will explore these issues in depth.

Children's search behaviour, the extent of adult involvement in this, and any resulting implications for the safe use of the Internet are also interesting and important themes. The latter may be particularly interesting considering both the area of the research that will examine policy initiatives and the extent the Internet appears, from the large scale NOP study, to be exposing children to material they would rather not see. The fact that the pilot study reported here did not find any widespread incidence of this only raises more questions and calls for more research. It may be that the problem only occurs when children are a little older than the ten to eleven year olds studied here; that the researcher simply did not probe deep enough, or that the users were, despite their tender ages, already sophisticated 'surfers' who could judge unpleasant material early enough to avoid it having an effect on them. Clearly, future research will benefit from a pragmatic and objective exploration of this important issue.

Previous research, and the study reported here, have shown that the Internet is being exploited by young people, that use has provided alternative sources for both school work and much pleasure away from the realm of formal education and that, in the limited pilot at least, little material adults may not wish their children to see appears to have been retrieved. The signs are hopeful – but further research is still much needed.

## References

1. DEE (Department of Education and Employment). *Connecting the learning society: national grid for learning*. (London: DFEE, 1977.)

2. BBC/LA (BBC and Library Association). *Libraries Project: promoting lifelong learning: project directory*. (London: Library Association, 1998.) (unpublished)

3. Cole, B. 'Racist propaganda hits the Internet on BNP website.' *Sunday Mercury*, 25 April 1999, p.2.

4. Robertson, L. 'Scott held in child porn raids.' *The Herald*, 20 March 1999, p.8.

5. Electronic Telegraph. *Police smash Internet child porn empire*. 3 September 1998. *http://www.telegraph.co.uk:80/et?ac=000923575723312&rtmo=a45TahuJ&atmo=777777Ft&pg=/et/98/9/3/npae03.html* (visited 3 February 1999)

6. Mead, D. 'An overview of the implications of the Internet for child welfare.' In: Carr, J. and Mullins, A. (eds.) *Children on the Internet: opportunities and hazards*. NCH Action for Children booklet. (London: NCH, 1998.)

7. Yow, D. 'Children First.' *Internet World* 7(4), April 1996, pp.92-93.

8. Leu, D.J., Diadiun D. and Len, K. *Teaching with the Internet: lessons from the classroom*. (Norwood, MA: Christopher-Gordon, 1997.)

9. Safford, B.R. 'The problem with the Internet: it is NOT the information highway.' *School Library Media Activities Monthly* 13(3), November 1996, pp.42-43.

10. Schacter, J., Chung, G.K.W.K. and Dorr, A. 'Children's Internet searching on complex problems: performance and process analyses.' *Journal of the American Society for Information Science* 49(9), July 1998, pp.840-849.

11. Martin, S. 'Internet use in the classroom – the impact of gender.' *Social Science Computer Review* 16(4), 1998, pp.411-418.

12. NOP (National Opinion Poll). *New European Internet Study reveals wide differences in usage across major markets*. 1998 *http://www.nopres.co.uk/internet/surveys/in08.htm*

13. Kafai, Y., Bates, M., Braxton, P.D., Childs, D., Ender, P.H., Lo, H.H., Martin, M., Rose, K. and Yarnall, L. *Building a foundation for information literacy: creating an annotated WWW-Index by children for children*. (Los Angeles: UCLA Graduate School of Education & Information Studies, 1996.) *http://www.gseis.ucla.edu/SNAP/report.html* (visited 19 May 1999)

14. NOP (National Opinion Poll). 'Over four out of ten British kids are now online.' *Internet upsets kids. http://nopres.co.uk/internet/surveys/pr22.htm* (visited 5 July 1999)

15. NOP (National Opinion Poll). *Internet upsets kids. http://nopres.co.uk/internet/surveys/pr20.htm* (visited 19 January 1999)

16. Newton-Dunnearly, T. '500,000 children upset by the Internet.' *The Mirror*, 8 December 1998, p.21.

17. Lawson, R. in conversation with Peter Williams, 10 February 1999. [Robert Lawson is a founder member of the NOP Internet team.]

18. OPCS (Office of Population Censuses and Surveys). *Ward and civil parish monitor: 1991 Census – Hertfordshire*. (London: OPCS, 1991.)

19. Marketer. 'Net user demographics: gender.' *Marketer*, 5 June 1999. *http://e-land.com/estats/demo_sex_2.html* (visited 5 June 1999)

20. Irwin, A. 'Chatting women move to Internet.' *The Daily Telegraph*, 1998. (FT Profile search)

## Acknowledgement

Many thanks to the staff and pupils of Bengeo Primary School in Hertford and in particular to Mrs Pledger and her class.

## Subscriptions

*Aslib Proceedings* is published ten times per annum and supplied to corporate members of Aslib. For 1999 the annual subscription rates are £170/$280 (£136/$224 to members of Aslib). For subscriptions and back issues please contact: Portland Press Ltd, Commerce Way, Whitehall Industrial Estate, Colchester CO2 8HP, United Kingdom. Tel: +44 (0) 1206 796 351. Fax: +44 (0) 1206 799 331.

Aslib, The Association for Information Management, is a world class corporate membership organisation with over 2000 members in some 70 countries. Aslib actively promotes best practice in the management of information resources. It lobbies on all aspects of the management of, and legislation concerning, information at local, national and international levels. Aslib provides consultancy and information services, professional development training, conferences, specialist recruitment, Internet products, and publishes primary and secondary journals, conference proceedings, directories and monographs.

Further information can be obtained from: Aslib, The Association for Information Management, Staple Hall, Stone House Court, London EC3A 7PB, UK. Tel: +44 (0) 20 7903 0000. Fax: +44 (0) 20 7903 0011. Email: aslib@aslib.co.uk *http://www.aslib.co.uk/*

Periodicals Postage Paid at Rahway NJ and at additional mailing offices.
**POSTMASTER:** Send address changes to PROCEEDINGS, c/o Mercury International, 365 Blair Road, Avenel NJ 07001

©1999 Aslib and contributors ISSN 0001 253X

ISSN 0001 253 X
Volume 51
Number 10
November/December 1999

23/2/2000

# Aslib Proceedings

New information perspectives

## Will online images prove to be a multimedia cornucopia for museums?

Edith Maier

---

## Access to information on the World Wide Web for blind and visually impaired people

Charles Oppenheim and Karen Selby

---

## Public libraries and EU information

Mike Cooper

---

## Using cluster analysis to determine the media agenda

Y.H. Kok, A. Goh and D.A. Holaday

Aslib THE ASSOCIATION FOR INFORMATION MANAGEMENT

*Aslib Proceedings* is a long-standing professional journal of international repute, which has a large and wide subscriber base. The journal brings currency, authority and accessibility to the reporting of current research, issues and debates in the broad area of information work. Above all the journal wishes to provide research and comment in a form that is easily and quickly understood: a fresh, rigorous, but unfussy, writing style is what is aimed for. The needs of busy practitioners (and academics) are very much in mind. Articles are referred to meet the need for accredited and authoritative information. To ensure that this does not result in an unacceptable loss of currency, articles are speedily refereed by the journal's large, eminent and multi-disciplinary Editorial Board (assembled for just such a purpose). In exceptional cases they may be referred to an external panel of consultants. The journal's monthly publishing pattern helps in maintaining the currency of reporting. In consequence the submission of research in progress articles is encouraged.

The boundaries of information work are rapidly changing and widening. *Aslib Proceedings* has a role to play in the respositioning of the information profession by encouraging submissions from information managers, providers, packagers and consumers of all kinds – and especially those from the related information fields of journalism, electronic publishing, communication and internet studies. An interdisciplinary Editorial Board has been assembled to assist in the process.

Prospective authors should submit articles of 2,000–8,000 words in length. Articles should be accompanied by a 100–150 word informative abstract and the numerical form of referencing should be adopted. In style and tone, submissions should fall between that of the journal's sister publications: *Managing Information* and *Journal of Documentation*. Articles should be furnished by email or on disc to David Nicholas, Editor, Aslib Proceedings, Department of Information Science, City University, Northampton Square, London EC1V 0HB. Email: nicky@is.city.ac.uk. The Editor is happy to provide advice on an informal basis to potential authors.

**Editorial Board**

Dr David Nicholas, *Head of Department, Department of Information Science, City University;* John Akeroyd, *Director, Learning Resources, South Bank University;* Sarah Blair, *Head of Publications, Aslib;* Peter Chapman, *Information Manager, NewQuest North East;* Professor Peter Cole, *Head of Department, School of Journalism, University of Central Lancashire;* Dr Tom Dobrowolski, *Senior Lecturer, Institute of Information Science and the Book, University of Warsaw;* Harry East, *Senior Research Fellow, Department of Information Science, City University;* Mike Hyman, *Managing Director, Head Software International;* Robin Hunt, *Creative Director, AreHause;* Nat Lievesley, *Network Manager, Centre for Policy on Ageing;* Helen Martin, *Information Manager, Research and Information Unit, The Guardian;* Dr Ian Rowlands, *Lecturer, Department of Information Science, City University;* Richard Withey, *Director, New Media, News International;* Ralph Adam, *Freelance Journalist.*

**Editorial Team**

*Editor:* David Nicholas
*Deputy Editor:* Ian Rowlands
*Editorial Manager:* Kay Nicholas
*Editorial Assistants:* Peter Williams (The American University), Colette Coles (London Research Centre)

# CONTENTS

325–334 Will online images prove to be a multimedia cornucopia for museums?
Edith Maier

335–345 Access to information on the World Wide Web for blind and visually impaired people
Charles Oppenheim and Karen Selby

346–360 Public libraries and EU information
Mike Cooper

361–371 Using cluster analysis to determine the media agenda
Y.H. Kok, A. Goh and D.A. Holaday

# Will online images prove to be a multimedia cornucopia for museums?

*Edith Maier*
*Garden Flat, 39 Holland Park, London W11 3RP*
*Email: e_maier@maier.vol.at*

**Abstract** *Faced with increasingly tight budgets many museums have come under pressure to explore new sources of income. The marketing of museum assets seems to offer one possibility especially since there is an apparently insatiable demand for content from the multimedia and 'edutainment' industries. Furthermore, providing access to the cultural heritage of Europe has come to be seen as a major aim of the EU and various programmes have been set up to encourage and facilitate this. Apart from the economic potential of digital museum content, the concomitant legal and security issues are investigated.*

## Introduction

This article reports the major findings, conclusions and recommendations of a study of the marketing of museum images which was commissioned by the Austrian Ministry of Education and Cultural Affairs (the Ministry) in the summer of 1998. It has to be set against a background of more autonomy for museums on the one hand and shrinking funds on the other, which is due to a recent change in legislation concerning the nine Bundesmuseen (Federal Museums) in Austria, among them the famous Kunsthistorisches Museum (KHM), art museums such as the Albertina, the Belvedere, the Museum of Modern Art Stiftung Ludwig (MMK-SLW), the Museum of Applied Arts Vienna (MAK), the Austrian National Library and other less well-known institutions, such as the Pathologisch-Anatomisches Museum with exhibits of human pathological specimens. Until recently the Federal Museums were administered and financed exclusively by various ministries, but from the beginning of 1999 onwards they will be gradually turned into corporations under public law and so became institutions largely independent of the Government and fully responsible for the management of their financial affairs. Subsidies have been frozen at their present level thus obliging museums to make up for the resulting shortfall either by making economies through greater efficiency and rationalisation and/or by tapping new sources of revenue or boosting existing ones.

The study was undertaken with the following objectives:

- to identify the status quo of asset management and marketing in the Austrian Federal Museums;
- to compare the Austrian situation with developments at European and international level;
- to identify, describe and evaluate different marketing approaches;
- to produce a business plan for the marketing of museum assets including the hardware and software requirements;
- to recommend how the museums could best proceed in marketing museum images.

The Project Team consisted of representatives from various museums, the Ministry and from a consulting firm which co-ordinated the Project and was responsible for conducting the research, summarising the results and producing recommendations.

## Austrian museums at the threshold of the digital age?

The status quo was examined through a series of questionnaires, interviews and consultations with museum staff and various experts, interlinked with discussion and analysis amongst the Project Team. What emerged was a very heterogeneous picture. On one end of the spectrum there are large museums with up to 300 staff, that have already digitalised parts of their collections and published them on the WWW, such as the KHM, the Albertina, the Ethnological Museum or the Museum of Modern Art. In fact the Museum of Modern Art was the first European art museum to go online with its own homepage in 1995. At the other end of the spectrum we find the small museums whose assets are less well known and therefore less commercially attractive and which lack both the human and the financial resources for carrying out large-scale digitalisation projects.

Although most respondents felt that the scanning in of digital images of museum collections would open up new opportunities and was, in principle, desirable, they lacked clear strategies about how to proceed. At the time of the study the total number of digital images from all the Federal Museums did not exceed 10,000, the bulk of which could be attributed to the Austrian National Library with its impressive collections of drawings and manuscripts. The museums' income from the trading of their assets ranged from virtually zero to £150,000. In 1997 the overall turnover of all the reproduction departments amounted to about £350,000.

Although the scientific and scholarly documenting and interpreting of a museum's collections tends to be a strong motive behind the scanning in of museum objects, providing educational and public access to a museum's collections is increasingly seen as desirable as well. All over the world, the educational community is interested in incorporating museum digital content into teaching and learning processes and the public at large has also shown a keen interest in virtual museums as is testified to by the high degree of popularity of museum web sites. The MMKSLW (*www.Austria.EU.net/mmkslw/*), for instance, receives 1,500 visits per day. One of the top addresses in the German-speaking museum world, the Art and Exhibition Hall of the German Federal Republic in Bonn (*www.kah-bonn.de/index.htm*), receives 2,000 visits a day.[1] According to surveys in the US museum web sites rank before the US Federal Government and the US Football Association.

The results from the questionnaire showed that the pricing and rating structures for reproductions varied greatly not only between museums but sometimes even between departments. In some museums, the situation was exacerbated by conflicting interests and clashes of responsibility among staff. Requests for reproductions came mostly from the scholarly and educational community; only a small percentage came from commercial clients such as advertising agencies, multimedia companies or graphic designers. Requests for digital images were extremely rare – the KHM, for instance, received only two such requests in 1997/98!

### *Attitudes towards online asset management and marketing*

Despite a lack of demand for digital images, there was a widespread feeling that Austrian museums should embrace new technologies so as not to 'miss the boat'. Besides, many museum workers have come to realise that familiarity and experience with Internet technologies might give them a competitive edge when it comes to attracting visitors. An increasing number of museums therefore have started to put information about their opening times and collections onto the Web, sometimes offering extensive documentation about exhibitions including previews as well as useful links to related sites or topics. Although they still have a long way to go before reaching the Internet-based museum experience of the ArtMuseum.net (*www.artmuseum.net*) presented by Intel Corporation, museums in Austria have been making great strides.

Few museums, however, have given much thought to the marketing of online images. Some have been approached by Corbis and other international picture agencies and are therefore aware of the possibilities of asset trading. Marketing in itself is something of an anathema for many curators. Only one of the museums, the MMKSLW, has a marketing department and advertises its holdings. But many museum workers remain distrustful of marketing and see it as contrary to or alien to a museum's real mission. Nevertheless, museums are starting to engage in fairly low-key promotional and public relations activities, but there is certainly no indication that marketing staff might begin to influence the curatorially defined mission.

## European and international context

Museums and galleries around the world are scanning in digital images of their collections and producing new products such as CD-ROMs, videos or web sites. The advent of digital museum content opens up new possibilities. People in remote places can pay virtual visits, they can view parts of collections that are in storage (typically about 50% to 98% of the whole collection) and museums can exchange documentation about their holdings among each other. It is generally recognised that digital collections will lead to new awareness of patterns in cultural development.

At the same time, the wider international perspective is also characterised by a parallel process of cuts in direct public funding, pushing museums into the marketplace, with a consequent need to emphasise income generation and marketing.

### *Museums in Britain*

The Team decided to go on a fact-finding mission to London in order to obtain an overview of asset management and marketing and gather examples of good practice. It was felt that in Britain museums had been exposed to market forces for longer and could therefore provide useful lessons. The visiting programme included five museums including the British Library, two commercial picture libraries, namely the Bridgeman Art Library and Corbis, as well as the consulting firm Pira UK.

The Bridgeman Art Library is – in its own words – 'the world's most comprehensive source of fine art images for publication, currently representing more than 750 collections around the world'. It has been in existence for over 25 years and recently launched its web site and published its catalogue on CD-ROM. Corbis was founded in 1989 by Bill Gates and has become the largest provider of digital images (over 1.3 million) by combining state-of-the-art technologies and photographic content, with only about 10% from fine art sources. Pira UK has published various studies about asset management and multimedia rights clearance, mostly on behalf of the European Commission.

As in Austria British museums depend either directly or indirectly on funding from central government. When successive British Conservative administrations sought to cut perceived waste in public funding, the museum sector had to develop new approaches to financial planning. Over the last two decades museums have been exploring new ways of raising revenue through retailing and trading, such as shops and mail order, merchandising, the rental of space for non-museum functions, fees for services (photographic, reproduction fees, lecture tours etc), subscriptions and the reclaiming of VAT. These days British museums earn up to a third of their revenues from these sources, something which compares favourably with the situation in the United States where – in 1988 – art museums raised between one-fifth to one-quarter of their revenues in this way[2]. Apart from the above mentioned sources British museums also rely on gifts, donations and membership, but not to the same extent as in North America.

As a result of the changes in attitudes and funding patterns, a new professionalism has developed in fundraising and income generation and there is hardly a British museum that does not have its own development and marketing department. It came therefore somewhat as a surprise to the Austrian visitors that their hosts should repeatedly emphasise the idea of the museum as a public good and that despite, or rather parallel to, the emphasis on cost-efficiency and value for money, a campaign for free access to museums had been waged since 1986. In Austria, national museums have been charging entrance fees, as has been the case with French museums. They do not or will not hesitate to increase them if they have to tighten their belts whereas the policy of charging at the national museums and galleries is very controversial in Britain, where up until the 1980s free admission was the rule. Today only the British Museum and the National Gallery do not levy charges on their visitors.

### *Presenting and marketing of museum assets*

British museums as a whole certainly proved to be more advanced when it came to espousing the new media such as the Internet and

CD-ROM for the presentation of their collections. Most of them have their own web sites and many have already transferred the whole or parts of their image libraries to the WWW, mostly in the form of thumbnail images.

The Natural History Museum (NHM), for instance, offers an excellent online catalogue with images of specimens and artwork from the museum's collections. Visitors to the site can print them out and download them for personal reference. Those wishing to make commercial use of the images in the NHM Picture Library still have to rely on large format transparencies, which can be borrowed from the Museum by paying a service fee. If they decide to make use of them, they have to make a written application for permission to publish. Other museums proceed or intend to proceed in a similar way but few make such a clear-cut distinction between educational and commercial clients (academic requests are referred straight to the appropriate departments).

Web sites and CD-ROMs are considered a useful addition to traditional marketing avenues because they make sourcing pictures easier: a web site or CD-ROM can be searched using all manner of fields and keywords including the name or nationality of the artist, the title or location of the work, period, style, or – in the case of the commercial agencies such as Corbis or the Bridgeman Art Library – picture content (e.g. jealousy, sunset). In the past (potential) clients had to rely exclusively on print media such as the directory published by BAPLA (British Association of Picture Libraries and Agencies), the *Handbook for Picture Researchers*, catalogues or information obtained at book fairs and other events. Most museum staff consider the credit-line an important and effective marketing instrument.

The Web is regarded primarily as a promotional tool that helps to advertise their collections and services, but not as a distribution medium. On the whole, museum staff were sceptical about demand for online images. Requests for digital assets tend to come from the press or TV stations. Traditional customers such as publishing houses still seem to lack the know-how or equipment. This contrasts with statements from Pira who predict a great increase in the demand for digital museum content. Nor does it coincide with the experience of Corbis, which – right from the start – focused on digital techniques and distribution mechanisms. According to Corbis demand for online images has grown from 5% to 50% in the last five years and over 25% of images are distributed online these days compared with 5% about five years ago. Visitors to the site can browse, search, preview, request and download images online. Images that are used in low-resolution form can be delivered over the Web immediately. High-resolution images can be delivered overnight via CD-ROM, or can be sent via Corbis' FTP (in the US) or ISDN (in Europe) service. Normally requests are dealt with within 24 hours.

Only in the case of royalty-free images, however, has it been possible to automate asset trading to such an extent as to make non-electronic contact redundant. As soon as rights handling is involved, human intervention is called for to negotiate licenses and provide quotations.

### *Marketing approaches*

Four marketing approaches could be distinguished:

1. in-house picture library;
2. non-exclusive cooperation with commercial agencies;
3. cooperation with non-profit locator agencies;
4. common museum platform.

#### *1. The 'do-it-yourself' approach*

Traditionally, museums have handled the management of assets in-house, usually by a staff member in the reproduction department. They tend to have developed their own policies and procedures for handling different types of requests that are sometimes systematised into a series of forms and standard contracts. In some museums, however, each request is evaluated on its own merits. The museum handles all correspondence and administration of reproduction requests and bears any associated costs and, in return, keeps all revenues.

Although the museums visited showed a preference for in-house solutions, BAPLA appears to have a harmonising effect on the operations of most picture libraries, both commercial and non-profit ones. It offers support and guidance to all its members (over 320 in July 1999), doing valuable work in the areas of intellectual property rights, price guidelines, technology, library management and standard contracts.

*2. Cooperation with commercial agencies*
If a museum does not have the expertise, resources or time to engage in activities such as digitalising and/or marketing ITS images, it can decide to have a third party handle the administration of reproduction requests or digitalise their holdings. Agencies offer a streamlined administrative process for users and also offer greater marketing potential because a large pool of content tends to be more attractive than an individual collection. Agencies are interested in purchasing the license to works from individuals or collections and then re-license them at a profit. Original rights holders receive either percentage royalties or are given a lump-sum payment for the unlimited re-use of their images over a specified period. If an agency does not own all the copyrights to the images, they usually supply the customer with the information required for rights clearance.

Terms and conditions and fee structures vary between agencies: the Bridgeman Art Library, for instance, charges an average commission fee of 50% and offers three year contracts; Corbis keeps up to 65% of the income and prefers ten year contracts. Contracts are non-exclusive in both cases, i.e. museums are free to offer their images through other avenues.

*3. Non-profit or locator services*
Users often find it difficult to locate images. Networked communications can offer a solution by creating a merged catalogue of images in museum collections while leaving control over the administration of rights in the hands of the rights holder. The Image Directory is a locator service founded by Academic Press in 1996. It provides access to fine arts images similar to stock photograph houses but without providing the image directly. It seeks only to distribute information about the objects but is not interested in buying rights to images, has no intention of becoming an image broker and will not authorize or distribute any reproductions of images. Requests are passed through to the holding institution. Such a model offers an attractive, one-stop shopping alternative, but it does not facilitate the rights clearing process. Users are still faced with a broad range of licenses and the terms and conditions of individual institutions.

The different approaches are not mutually exclusive, but can complement each other. A museum may have its own picture library but decide to market part of its collection through a commercial agency so as to reach the American market or a different group of clients. They might also place their images with a locator service and thus increase their accessibility.

*4. Common museum platform*
Another approach consists of presenting and marketing a museum's images through a common museum platform. Britain does not have a common cultural heritage platform as, for example, the Canadian CHIN or the Australian AMOL platforms, which is why information on the fourth option was gathered mostly through interviews with experts, publications and Internet research.

The first systematic museum network was the aforementioned Canadian Heritage Information Network (CHIN), founded in 1972. Its fee-paying members receive support with technical, legal and security issues; they are granted access to various databases and bibliographic information as well as to standard licensing contracts.

The Art Museum Image Consortium (AMICO) was launched in 1997 and is a non-profit association of art collecting institutions working together to enable educational use of their digital documentation. It aspires to create a vast art database that eventually may exceed 250,000 images (at present just text and image, but over time also sound and moving image) compiled by its dues-paying members, which include the Metropolitan Museum of Art in New York and the Art Institute of Chicago. Unlike online art exhibits, which are aimed at general audiences, the AMICO Library is a cross-museum database that will function best as a research tool sharing the documentation in museums.

In Europe ArtWeb is pursuing similar aims, but has not gone 'live' yet. The project, which receives grants from the European Commission's Info 2000 scheme, brings together museums such as the Réunion des Musées Nationaux, Paris and commercial picture agencies such as the Bridgeman Art Library to produce a joint online service of over 200,000 images from over 1,400 museums, photographers, private collections and individual artists. It will allow clients to conduct multilingual research, select, order and download images 24 hours a day.

**Recommendations**

The impressions gained from London, together with the results from interviews with experts and desk research, led to the following recommendations:

- create structures which promote and encourage exchange and cooperation among museums;
- adopt common technical, legal and organisational standards;
- establish price guidelines so as to harmonise pricing within and between museums;
- adopt a pragmatic approach to digitalisation rather than look for the perfect solution;
- depending on museum resources, outsourcing certain work such as scanning/digitalisation or marketing is advised.

***Don't go it alone***

Sharing expertise and adopting common standards and guidelines were the key recommendations. Various initiatives are already underway to respond to this widely perceived need for more cooperation. In Europe, the Memorandum of Understanding (MoU), a joint initiative by the European Commission's Directorates-General XIII and X, sets out to create conditions for harmonious and balanced development of the market of multimedia services and applications to access the European cultural heritage. It was launched in 1996 and has since been signed by more than 450 organisations (not only from Europe). The aim is to establish a cooperative framework among museums and industry and explore de facto and de jure technical standards as well as the issue of protecting intellectual property rights, which is considered crucial for the development of an open market. Although the MoU has been generally welcomed by the museum sector, some feel that in addition to understanding, more commitment is called for to meet these challenging goals.

***Common standards***

Developing and then adhering to common standards is a sine qua non for closer cooperation. Standards are necessary in several fields:

- description, content and presentation standards[3];
- legal standards;
- pricing and fee structures.

*Description, content and presentation standards*

There are hardly any formal standards available and in many cases no specifications which are sufficiently widespread to be considered de facto standards. However, in most areas standards are evolving, e.g. RDF (Resource Description Framework) under development by the W3C (World Wide Web Consortium), which is expected to provide a general-purpose metadata syntax. XML (Extensible Markup Language) is another emerging standard, which offers a kind of compromise between too much generality (SGML) and too much specificity (HTML). SMIL (Synchronised Multimedia Integration Language), an application of the XML syntax, is of great relevance and potential for museum applications because it concerns the delivery of real-time (streamed) audio-visual content.

When it comes to image formats which define ways of capturing, storing and compressing digital images, we are faced with several formats. GIF (Graphics Interchange Format) and ISO JPEG are the de facto standards, in particular on the Internet. They are currently the safest choices when targeting the Internet environment. The Tag Image File Format (TIFF), developed by Aldus and Microsoft, and the Photo Compact Disc

(Photo CD) format are also commonly used. Another widerspread format is Microsoft's BMP. A new specification by W3C, called PNG (Portable Network Graphics), aims at replacing GIF in the medium to long term. The images are usually compressed using JPEG.

Levels of resolution and compression used in specific applications tend to vary according to the hardware and software used and the purpose of a particular application. For instance does a museum want to create images for archival or preservation purposes? Does it want to reach the scientific or the educational community or the general public? Are the images to be published on the WWW? Experts advise that any project should ensure that the type of image format used fits the purpose for which it is intended.

### *Rights protection and clearance – towards a 'one-stop-shop'?*

Many museums are reluctant to distribute their documents and images in public networked environments because of the ease of intercepting, copying and redistributing electronic data in their exact original form. They are therefore concerned about the protection of intellectual property rights and the moral rights of artists whose works are in their care. Encryption is one method of protecting data against unauthorised access. In the case of images, digital watermarking is gaining ground because it allows their spreading to be tracked. A digital watermark is an identification code giving information about the copyright owner, the creator of the work etc. and is permanently embedded into digital data. There are no de jure or de facto standards as yet, but a series of proprietary solutions (e.g. those distributed by Digimark or Highwater). However, the IMPRIMATUR project (ESPRIT) has produced a comprehensive set of requirements on watermarking and represents a centre of expertise for digital property protection of cultural content.

Since the images on most museum web sites tend to be low-resolution thumbnails (usually 72 dpi) museums need not worry about potential abuses since these are quite useless for commercial purposes. Should they decide to offer high-resolution images for downloading at a later stage digital watermarks can provide efficient copyright protection.

It is not the security issue that acts as a barrier to the online trading of images but rights administration and clearance. Readily identifying and gaining rights to use museum digital documentation can be a time-consuming and frustrating task. Contemporary art, in particular, presents a host of copyright concerns, which is why modern works tend to account for only a small percentage of the content in commercial picture libraries. Not surprisingly, multimedia producers have been pushing for 'one-stop-shops', i.e. organisations which not only provide information, act as intermediaries and manage rights (with fixed fees) but also deliver the works needed for production. Both in Germany and Switzerland central clearing organisations have been founded in order to streamline the rights administration process. From a client's point of view this has the advantage that a large amount of content from differing collections can be made available at a single point. At present, however, the process still takes a long time – 45 days in the case of the Swiss clearing centre, which is apparently due to the delays incurred by dealing with rights holders in other countries. Many museums hesitate to assign rights to third parties because they then have little control over how and where an image may be used. This is why many prefer information or service centres where relevant information is accessible on how and where to clear permission for reproduction.

### *Pricing and fee structures*

As already mentioned, BAPLA issues price guidelines that museums in Britain tend to follow. Austria lacks such an unifying institution and prices and fees fluctuate widely, not only among museums, but at times within a single museum. Since the importance of a common pricing policy was realized before this study was commissioned, an informal working group has been set up with representatives from different museums with the explicit aim of harmonizing pricing policy.

### ***Pragmatic approach to digitalisation***

Creating a large digital image archive poses a number of technical challenges with regards to capturing, processing, storing, delivering, and managing images. The term 'archive' implies that all digitalised images are not only optimized for current work flows and imaging devices, but also maintain usability on future, as yet unknown, delivery and output systems. Both museum staff and experts warn against looking for the perfect solution but recommend a pragmatic approach: digitalise images according to demand from outside clients, i.e. the commercially successful images, and at the same time develop a long-term strategy for digitalising the whole collection. They also warn against investing too much into the latest technologies or state-of-the-art equipment because technologies are changing so rapidly.

Choosing the correct scan resolution is dependent on the purpose of the archive. Although it would be ideal from an archival point of view to digitalise at a level that captures all the information content in an image, in reality the necessary scan resolutions (up to 4,000 pixels per square inch for modern film material) are hard to achieve with current CCD scanners for formats larger than 35 mm. The resulting file sizes are huge, up to 960 Mb uncompressed for a 4 × 5 inch original. A few museums such as the National Gallery or the Uffizi in Florence have started to digitalise their collections at very high resolution (at 1.4 Gb per square meter or about 300 Mb per image). Most museums, however, scan in their works at approximately 30 Mb to 40 Mb per image.

For browsing on the Internet the size of image files should not exceed 100 kb – at least at present transmission rates – to make it easy to search through the database and keep the image retrieval time to a minimum. Picture archives which provide web access to images tend to store three resolutions of access files online: a thumbnail (about 140 × 110 pixels), a preview file and a larger preview file and/or the possibility to zoom in on an image. Pixel dimensions range between 256 pixels in the case of Corbis (height or width of the image), or 600 × 500 pixels (Marburg Archive) for preview files and 480 by 640 pixels for the enlarged files (Corbis) compared with 1400 × 1100 pixels used by the Marburg Archive.

The archive images are normally stored in uncompressed RGB or Lab TIFF format. Lossless compression is achieved by hardware compression when writing the tapes. Proprietary file formats for archive files should be avoided. Theoretically, an archive file might not be accessed until the next time the archive is recopied onto the next generation storage medium, which might occur in seven to ten years. At that point source code to interpret the image file format needs to be readily available.

Improvements in software, in the storage capacities of servers, in Internet transmission speeds and in desktop computer processing and display capabilities will, no doubt, facilitate the transmission of high-quality digital images over the Internet. At present, downloading whole files is still too cumbersome for the casual user. Software in this area is rapidly developing and there are products on the market that are already used quite effectively with museum materials. An interesting product can be found at the Zoom Project web site maintained by the Fine Arts Museums of San Francisco in collaboration with the University of Berkeley. Its ImageBase (*www.thinker.org/imagebase/index.html*) utilizes the Grid Pix software suite which allows users to scroll within a large image to view portions at different levels of magnification.

There is consensus that, in the long run, an object-oriented approach complemented with a sound metadata syntax, which allows deep searching and retrieval of images and data, is the most promising approach to adopt for setting up digital image archives.

## Conclusions

After evaluating the different approaches to marketing museum assets the study concludes that a common museum platform or a consortium of museums that collectively presents and promotes the pooled content of its members would be the most effective means of making digital information accessible to the public.

Acting consortially would have the following benefits:

- building up a collective library of art from Austrian museums;

- exerting influence with vendors to reduce costs of capturing the data;
- enabling its members to have access to collective funding;
- negotiating digital rights with artists and artists' estates and with museums in other countries; and finally
- collective licensing and distributing of digital images including billing and fee collection.

However, three major prerequisites have to be met before such a platform can operate. They are:

1. a critical mass of digital content;
2. effective intellectual property rights clearance mechanisms;
3. adequate standards.

Austria still has a long way to go before it can fulfil all these prerequisites. It is lacking in all three areas. Digital content is growing but is still very paltry compared to commercial picture libraries or museum platforms such as AMICO; there is no indication as yet that rights clearance is becoming any quicker or less time-consuming; and the adoption of common adequate standards has not gone beyond the discussion stage.

To remedy this situation the study suggests adopting a gradual approach and – as a first step – setting up a service and information centre. Apart from giving advice about documentation practices, information infrastructures, digitalisation, electronic collection management and technical and legal matters, such a centre would in cooperation with the existing working group (cf. 'Common Standards') issue price guidelines and recommend standards. It could also offer consultancy services with regard to fundraising, marketing and museum management. Rather like the German digitalisation centres in Munich and Göttingen it could offer practical services that might be particularly attractive to smaller museums, such as digitalising collections and storing the digital images. This would require additional investment in hardware and software for the centre, but the savings achieved through economy of scale might outweigh the attractions of the DIY approach. At a later stage, and if its members so wish, the centre could extend its activities to marketing and licensing the images including delivery and fee collection.

It is reckoned that an investment of approximately £200,000 would be sufficient to launch a service and information centre. Most other museum platforms are funded through membership fees or subscription and/or by regional or state governments. As the Ministry seems reluctant to commit itself financially, private sponsors would have to be found or public-private partnerships formed to kick-start such an enterprise which – according to preliminary figures – could yield an operating profit after about four years. Unfortunately, recent changes have actually reduced the Ministry's ability to enforce cooperative structures and it looks as if the larger museums, in particular, have decided to go it alone. They do not rule out cooperation altogether but show scant interest in founding a consortium, although some of them do envisage offering digitalisation services to other museums so as to make better use of their equipment.

### *Economic potential of digital museum content*

Whether the study's suggestions are taken up or not, museums will have to look beyond their traditional structures towards the appointment of marketing managers and directors to exploit their collections so as to raise money. The question is whether asset marketing could develop into an important source of income – this, after all, was the question the study set out to answer.

First of all it has to be stressed that digitalisation alone has no significant artistic, scientific or economic value. It is the packaging of content which is crucial. Only by adding knowledge and presenting digital content in an appealing, accessible and understandable form does content develop its educational and economic potential. Some museums have realised this and have developed user-friendly, accessible and interactive environments which encourage individual and group exploration and learning.

There is no doubt that companies such as Corbis are doing good business and that the multimedia industry has a keen interest in fine art content. For the time being, however, the potential revenue museums can expect from marketing their assets should not be overestimated. Even in the case of an efficiently and professionally run in-house picture library, which only deals with commercial clients, the profits do not amount to more than 1–2% of the revenue raised by the museum from sources other than the public purse. However, since most museums have started to digitalise their collections anyway, be it for scientific, archival or promotional purposes, they might as well raise income from their digital images using a mix of the marketing strategies, such as in-house picture libraries, non-exclusive cooperation with commercial agencies and/or non-profit locator agencies. But in order to be able to compete with big players such as Corbis in the thriving multimedia market, museums would have to form alliances, streamline their procedures and pool their resources. They might already have missed the boat in this respect.

### References

1. 'Surfer-Paradies. Vergleich: Museen im Internet.' In: *Kunstzeitung* 35, July 1999, p.1.

2. Quoted in: Dickenson, V. 'The economics of museum admission charges.' In: *Museum Management* (London/New York: Routledge, 1994, p.107.)

3. See also Frey F. 'Digital imaging for photographic collections: foundations for technical standards.' In: *RLG DigiNews* 1(3), 15 December 1997. *http://www.rlg.org/preserv/diginews/diginews3.html#com.*

### Selection of URLs relevant to this article

Web site of the Austrian Federal Museums
*www.nhm-wien.ac.at/BundesMuseen/MUSEEN.htm*

Bridgeman Art Library
*www.bridgeman.co.uk*

Corbis
*www.corbis.com*

AMICO – Art Museum Image Consortium
*www.amn.org/AMICO/*

AMOL – Australian Museums Online
*www.amol.phm.gov.au/AMOL/MuseumResources/*

ArtWeb
*www2.echo.lu/info2000/en/mmprojects/artweb.html*

BAPLA – British Association of Picture Libraries and Agencies
*www.bapla.org.uk*

Bildarchiv Foto Marburg
*www.fotomr.uni-marburg.de/index.htm*

Canadian Heritage Information Network
*www.chin.gc.ca*

Image Directory
*www.imagedir.com*

***URLs and link collections of general interest to museums and picture libraries***

The World Wide Web Virtual Library: Museums
*www.icom.org/vlmp/*

Extensive link collection under:
*www.artnetweb.com/resource/resource.html*

ICOM – International Council of Museums
*www.chin.gc.ca/icom*

CIDOC – International Committee of ICOM for museum documentation
*www.cidoc.icom.org*

Getty Conservation Institute
*www.getty.edu/gci*

Getty Information Institute including access to the following databases:
  The Art & Architecture Thesaurus Browser
  The Union List of Artist Names Browser
  Getty Thesaurus of Geographic Names
*www.ahip.getty.edu*

# Access to information on the World Wide Web for blind and visually impaired people

*Charles Oppenheim and Karen Selby*

*Department of Information Science, Loughborough University, Loughborough, Leics. LE11 3TU*
*Email: c.oppenheim@lboro.ac.uk*

**Abstract** *The Internet gives access for blind and visually impaired users to previously unobtainable information via Braille or speech synthesis interpretation. This paper looks at how three search engines, AltaVista, Yahoo! and Infoseek presented their information to a small group of visually impaired and blind users and how accessible individual Internet pages are. Two participants had varying levels of partial sight and two Subjects were blind and solely reliant on speech synthesis output. Subjects were asked for feedback on interface design at various stages of their search and any problems they encountered were noted. The barriers to access that were found appear to come about by lack of knowledge and thought by the page designers themselves. An accessible page does not have to be dull. By adhering to simple guidelines, visually impaired users would be able to access information more effectively than would otherwise be possible. Visually disabled people would also have the same opportunity to access knowledge as their sighted colleagues.*

## Introduction

Since the Internet was introduced in 1969, it has grown rapidly and is expected to continue to do so over the next few years[1]. There are already over 50 million Internet pages, and more than 20,000 Internet sites are added each day, with each site possibly having hundreds of pages[2]. The major driver of this growth has been the World Wide Web (WWW). There are several possible reasons for the growth of the WWW. These include: ease of publishing; low cost of owning a web site; ease of updating; and the potential to reach a wide audience[3].

There is one group of users, however, that have been largely ignored in the rush to use the WWW. These are blind or visually impaired users, who have particular problems in accessing web material. Whilst accessibility issues to sighted users may be a matter of response time, getting lost in cyberspace, or poor performance of search engine retrieval, these are issues that may be inconvenient, but they are not insurmountable problems. The considerations are not the same for a blind or visually impaired user trying to access the same information.

Problems of accessibility to a person who is visually impaired covers all those for sighted users, plus a number that are unique to this group of users. These include the issue of screen design, the use of font size, colour, the use of patterns in screen backgrounds that make the text difficult to read and an excess of graphics. These features, designed to be appealing to the sighted user, may make Internet pages inaccessible to a visually impaired user.

A person can register as blind in the UK if they can only see the top letter of the eye test chart (used by opticians and doctors) or less, at a distance of three metres. This is sometimes called 3/60 vision: the blind person can see at 3 metres distance what a person with normal vision can see at 60 metres. So the term 'blind' does not necessarily mean that a person cannot see anything at all[4]. Therefore, some people who are registered blind may be able, with difficulty, to see objects placed directly in front of them three metres away.

A person may register as partially sighted in the UK if they can only read the top letter of the eye test chart at a distance of six metres. If they can read the next three lines down but their field of vision is limited they may still qualify for registration. A partially-sighted person may not be able to see the letters on the eye test chart any better than 'blind' people, but the field of vision may be less limited, or they may be able to see more clearly than a blind person but only within a very limited area. Partially sighted people can see less than sighted people can. Objects placed in front of the person at a distance of six metres may be seen with some difficulty[4,5].

In the UK, the number of registered blind and visually impaired people is set to rise over the next twenty years due to population increase[6]. The total current population is approximately 58.4 million people. For people who are over sixteen years of age and have a visual impairment:

- 1.1 million are eligible to be registered blind or partially sighted (not all people who are eligible register);
- 4% of blind people have no light perception;
- 1.7 million people are unable to read standard print;
- 36% of blind people can read large print;
- 2% (approximately 19,000) of blind people are fluent Braille readers;
- 30% of blind and partially sighted people use audiotape for their information needs[7].

The USA has an estimated visually impaired population of 8.6 million (3.4% of the total population). The number of American people who qualify as legally blind is approximately 580,000 and a further 1.8 million people have severe visual impairment[8].

People with disabilities represent a significant number of users for Internet applications[9]. The ability to communicate and collect information from anywhere in the world from the home has expanded the opportunities of disabled people to participate in society more effectively than ever before[10]. At present, most visually impaired people have to depend on radio, television or information which is read to them either on cassette or by reader, and cannot readily go out and buy a newspaper, magazine or browse through the bookshelves[11]. The WWW can provide information and services not otherwise available for the visually impaired, if the Internet page design is accessible to them. For disabled users there may be time constraints as they tend to take longer to find the required information[11]. The WWW is able to provide visually impaired people with a means to access information. The information can be used by the persons themselves, their families, teachers and employers[9]. For this reason the Internet has the power to change the lives of disabled people[12]. In addition, intranets in the business sector should be accessible to all members of the workforce, including disabled users[3].

## Accessibility issues

The design of web pages is an evolving technique for which there has been no set standard. HCI design principles have been established for general release software and these can be interpreted for web design techniques. Shneiderman states that there are eight rules of interface design[13]:

1. strive for consistency;
2. enable frequent users to use shortcuts;
3. offer immediate feedback;
4. design dialogs to yield closure;
5. offer error prevention and simple error handling;
6. permit easy reversal of actions;
7. support internal locus of control;
8. reduce short-term memory load.

These design principles apply equally to sighted and visually disabled persons. With any user-interface design process, the starting point must be: who are the users? The WWW is a global information resource and a web page may be accessed by a diverse range of people from different cultures and with differing abilities. Web pages can be created in various formats, e.g. text only, to provide access to disabled users[14].

It is difficult for most sighted users and designers to understand the problems faced by blind users. 'When blind people use the Internet, ... it is like trying to use Netscape with your monitor off and the mouse unplugged....'[15]. This poses unique problems when faced with a web site that relies on graphical interaction with the users[16]. The World Wide Web Consortium (W3C) is responsible for encouraging standards on the Web and launched the Web Accessibility Initiative (WAI) in 1997[17] to identify the problems facing disabled users and provide guidelines for designers who want to make their pages accessible to 'all' people. The needs of blind and visually impaired people appear to be overlooked in many cases of Internet page design.

**What the law has to say**
Recent Federal legislation in the United States, primarily Section 508 of PL (Public Law) 99–506 and the Americans with Disabilities Act 1990 (ADA), addresses accessibility problems faced by persons with disabilities in the workplace and community[18]. The Act requires that the Government and business sector provide an accessible medium for communication whenever they communicate through the WWW[19]. In the UK, The Disability Discrimination Act 1995 Part III, outlines specifically the Rights of Access to Goods, Facilities, Services and Premises. This Act is due to be amended late in 1999 to ensure that service providers are enabling disabled users to access their facilities[20]. The problem then ensues as to whether an Internet Service Provider is responsible in this respect for the content of the web pages it carries. It is likely that general principles that apply to other legal areas, such as defamation and copyright infringement, will apply in this case, namely that an ISP is responsible for the content it offers so long as it is aware of that content (which includes having its attention drawn to some issue) and so long as it is able to exercise control over whether that content appears or not[21]. There is no single ruling body for the Web, although there is voluntary co-operation[22].

**Page design issues**
The increasing popularity of the WWW causes unique problems for visually impaired users, and the features that provide elegance for some users, present barriers to others[23]. To enable access to visually impaired users, the design of the web page has to be 'screen reader friendly', and to aid the developers in this aim, they should endeavour to view their web sites with a text only browser[9]. There are guidelines and standards on accessible web page design, most notably the Web Accessibility Initiative (WAI). Other organisations concerned with accessibility issues include:

- The TRACE Center. TRACE is a non-profit research centre that focuses on making off the shelf technologies and systems like computers and the Internet, more accessible for everyone through the process known as universal, or accessible design[24].
- The Royal National Institute for the Blind (RNIB) who provide a wide range of practical solutions to common design related problems[25].
- The Yuri Rubinsky Insight Foundation WebABLE project provides web page designers with useful examples of accessible and inaccessible code for easy comparison. Following these instructions is relatively simple and shows that with some forethought into web page design, accessibility is not difficult to implement[26].
- Microsoft also publishes guides for accessible design and a Powerpoint presentation that is available to download from their web site which outlines the most common mistakes that are made when creating inaccessible web pages[27].

The major features that are of importance in page design for those with visual disabilities are:

- Use of HTML code standards, as extensions to the HTML standard cannot be validated, and makes Braille translation very difficult. Where extensions need to be used, an alternative presentation should be offered to the user.
- Keep page design simple and consistent throughout the site.
- Navigation throughout the site should be simple, consistent and clear.
- A text index or site map should be available to direct users to relevant pages.
- Keep graphics to a minimum.
- All images should be accompanied by alternative text (ALT-TEXT), which describes the image in a short sentence to enable Braille or speech synthesis to relate to the user the content of that image.
- Where images are used for eye catching purposes only or they are not important to the information content of a page, the ALT tag should be set to null (ALT="") to allow the browser to miss out the image.
- If image maps are used, there should be an alternative method of selecting options provided such as a text only alternative.

- Links should be described in ALT tags to make it clear to where it will take the user. Some access devices relate links out of context, so each link should have a clear title, 'Click Here' or 'Go' is not adequate.
- Links should not appear directly next to each other, as some access devices will interpret a group of links as one single link. Use separate lines for each link or separate the links with some text if possible.
- Do not use <BLINK> tag as this renders Braille and speech synthesis inoperative.
- Provide an alternative text (ALT tag) for all Java applets. It is frustrating to a visually impaired user to not know what is happening on screen.
- Use contrasting font colour and background colour. Do not use patterned or textured background designs, a solid colour is best.
- Ensure that your chosen colour scheme may be over-ridden by the browser settings if necessary, as some users may have their own preferences.
- Avoid the use of italic text, as this can appear 'wobbly' on screen.
- Avoid underlining text, as this may be difficult for some users to read.
- Do not use a font size that is too small or too large. Allow the user to set their own preferences by using browser settings.
- Capitalisation of words is difficult to read and should be avoided.
- Avoid the use of frames on a site, as some access devices cannot read the information held in frames. If frames must be used, a text alternative should be provided. Home pages should not use frames, as a user will have difficulty locating the text only version if the link is within a frame.
- Tables should be used with care as some access devices may interpret them line by line, which is confusing to a user.
- Provide text description for all movie clips or animation. This will enable the user to gain as much information from the text as sighted users get from viewing the screen.

To implement these guidelines, designers need to be more aware of the needs of the minority of Internet users. Implementing these guidelines does not mean that a page has to be dull, it just requires more forethought to achieve the same, or a very similar effect. Approved standards for accessibility are available on the WWW, for example 'Bobby' a Web Site Accessibility Verifier[28]. This is a freely available to download application from the Centre for Applied Special Technology (CAST)[29]. The application identifies HTML compatibility problems that may prevent pages from displaying correctly on different web browsers. The analysis of accessibility is based on the WAI Page Author Guidelines and all pages on a web site must meet these requirements in order to display the logo. An individual page can also be analysed and will be approved if it contains no accessibility errors. The errors that have been identified are reported back to the designer in the form of a comprehensive summary, the designer can then reassess the site and make corrections. Bobby is unable to analyse JavaScript for accessibility and it does not accept HTTP cookies. The Bobby checker enables novice designers to ensure their pages are accessible. The feedback from individual pages is comprehensive and simple to understand.

Another accessibility symbol is available from the National Center for Accessible Media (NCAM)[30]. An early goal of its Web Access Project was to develop a generic, public-domain symbol to indicate that a web site was designed with accessibility features. The symbol does not guarantee the page is 100% accessible, or that it was designed following WAI guidelines. Use of the symbol is at the designer's discretion.

In addition, guidelines for companies who wish to use the Internet as a source of information have been published[17]. There are effective corrective measures that allow access by all the users to all the information[31]. There are also documents that describe the problems and solutions available on the WWW[32].

Nonetheless, a survey in 1998 of 78 UK web sites that were specifically aimed at disabled people using the 'Bobby' checker showed 12 that were totally inaccessible, 44 with considerable accessibility problems and only 12 sites rated highly for their accessibility[33].

## Methodology

We undertook a study using four blind or visually impaired students to assess their reactions to a number of Internet search engines. The four Subjects chosen for this research were attending a foundation course within the RNIB College that includes mathematics, English and computer skills on the curriculum. They represent a wide range of visual impairments. Students were identified by their lecturer as having a range of difficulties. Their sight impairment ranged from:

- partially sighted but could use the computer interface with some difficulty;
- partially sighted and required an eye magnifier to use the computer interface;
- blind with very limited visual feedback from the screen and also reliant on speech synthesis;
- blind and solely relied on speech synthesis for text conversion.

The students were all mature students. They were all very willing to co-operate with the study and enthusiastic to convey their views on particular points that arose. They were all relatively new to computing and using the WWW. The research was carried out using the RNIB facilities. At the time this research was carried out (autumn 1998), students with partial vision were introduced to the WWW using Yahoo! as the default search engine and students who were dependent on speech used pwWeb Speak, which defaults to the AltaVista search engine. The RNIB College used Netscape Navigator 3.0 and pwWebSpeak (version 1.4.3). pwWebSpeak is a browser designed for users that wish to access information and services in a non-visual manner. pwWebSpeak is designed specifically to interact directly with the information on the Internet pages and to translate the information content into speech. There are many shortcut keys incorporated into the software that enable the user to carry out a variety of tasks such as reading the text sequentially or reading the hypertext links only[34].

We asked that each Subject accessed a search engine interface that they were familiar with, and two search engines that they had not previously encountered. The Subjects were interviewed before the experiment took place to ascertain their prior knowledge and their preferred search engine. As each Subject had similar Internet experience, the same three search engines (AltaVista, Yahoo! and Infoseek) were used in this research. The same search was given to all the Subjects. The topic related to the rise of recorded music that is available on the WWW, and the corresponding issues that arise from this. Two tasks were set: a general question on the broad issues, and a question on the predicted growth of music sales on the WWW for the next few years (the time scale was undetermined). Once the Subjects were clear about the questions being asked, they proceeded to search the Internet.

When a new interface was displayed or when speech synthesis was stopped, the Subjects were asked questions on their impressions of an interface. The questionnaire was filled in by one of the researchers and included a combination of open-ended questions and ranking scale questions. Often it was found that the Subjects were stating their preferences without any prompting from the researchers because of their enthusiasm to relate the difficulties they encountered.

To carry out the searches, we used the search engines, AltaVista, Infoseek and Yahoo!. A text version of AltaVista[35,36] is available, and can be clicked on at the bottom of the page by using the 'Text-Only Version' hyperlink.

## Results

The results are presented in search engine order. Subject One was partially sighted; Subject Two was registered blind, but retains a little vision; Subject Three was registered blind but can see objects at a distance; Subject Four was totally blind. All were mature male students.

### *AltaVista*

Subject One thought the design of the entry screen was good, particularly the use of yellow to highlight the entry field for search criteria. However, he encountered problems with accessing text information. Generally, the font size was too small to be read comfortably, though the colour and background colour had a good contrast. The highlighting of the entry field enabled him to immediately enter the search box. Therefore, the searcher did not consider the option of using the categorised information search. When the searcher was asked to appraise the interface as a whole, problems with the order of the text layout were reported. The title of 'AltaVista Highlights' was not aligned, which meant the Subject had to visibly search for the section heading for those hypertext links. This was unsatisfactory to the searcher. Graphical information was displayed on screen, and although the searcher was aware of the presence of the image, in order to read the information in advertising banners the searcher had to move closer to the screen to read their content. Until the text within a graphic had been accessed and identified to be an advertisement, it could not be ruled out as being unimportant and may have been information to aid their search. Overall, AltaVista was not well received by Subject One.

Subject Two found the use of colour to be beneficial. The text was too small, but the colours used provided a good contrast between text and the background. Advertisements required close investigation before they could be identified as advertisements. This required physically moving around the screen in order to access the information. Subject Two appreciated the highlighted search box, and rated AltaVista the best of the search engines.

Subject Three used PwWebSpeak. As a search is entered, feedback is given to the searcher, letter by letter including any spaces entered, to ensure that the correct letters are used. When the search term is entered the searcher is given feedback on what is happening by phrases such as 'Requesting page', 'Working' and 'Sending request' When the page is being downloaded the searcher is informed of 'Receiving page', along with the number of bytes received. Once the page has been downloaded the message 'Preparing page' and 'Page is ready' is given. The Subject found this speech feedback direct and useful. The speech software then however reported:

> 'Comma link comma AltaVista search service; a form for you to complete; descriptive label ask AltaVista a question; or enter a few words in; select value from list; enter value for entry field; select from radio button list change; options; submit button submit; search dash; comma link comma advanced; dash; comma link comma Usenet; enter form; AltaVista found four hundred and eleven thousand six hundred forty result pages for you; one comma link comma two comma link comma three comma link comma four comma link comma five.'

The Subject found this report back confusing, as it gave parts of a sentence at a time. The searcher did not understand the 'one comma link' etc, and had no idea that these were links to the pages of search results. The searcher lost interest while these links were being related back. He decided to access one of these links, and received the following verbal message:

> 'Comma link comma AltaVista search service; a form for you to complete; descriptive label ask AltaVista a question; or enter a few words in; select value from list; enter value for entry field; select from radio button list change; options; submit button submit; search dash; comma link comma advanced; dash; comma link comma Usenet; enter form; AltaVista found four hundred eleven thousand six hundred forty Web pages for you'.

He found this speech to be frustrating and irrelevant. The searcher also pointed out that skipping this text was not viable, as there is no indication of where the repetition ends. At this point the searcher became confused, and after listening to this speech again a link was made to another results page, and once again the above text was related back. This was because the searcher had only accessed the top part of the page, and the actual links to Internet pages appeared underneath the result page count, which had not yet been accessed. After repeating accessing the result pages once more, the searcher decided to carry on down a page where the information that was required could eventually be found. The searcher found this to be very frustrating, as the result pages appeared before the

Internet site links. Therefore the information that was required did not appear until listening to irrelevant text. There was also no indication of how much text there was to read, so the searcher did not know if the entire page had been accessed already, or only part of it.

The Subject also found it confusing that the link is found on the text, i.e. the title of the page, and not on the URL (pronounced 'earl'). The speech output gives the URL address letter by letter including 'period' for 'dot', and 'dash' for 'slash'. It also announces the date of last modification, e.g. 'last modified September twenty fourth nineteen ninety eight dash page size nineteen dash in English; comma link comma translate one hundred and ninety two'. The searcher pointed out that repeatedly listening through URL's announced letter by letter, and date of last modification was both confusing and time wasting, and of no help in the search.

Subject Four encountered similar speech output to Subject Three, but knew that the results were below the page links listing. Nonetheless, he could not miss out the starting speech because he did not know where the repeated text finished and the actual Internet page results began. He did not understand the page links and did not understand why they were there. Interpretation of the speech synthesis was sometimes very difficult. He considered date of last modification and byte size to be irrelevant. The searcher also became disorientated by the duplication of hits. He quickly became fatigued.

**Yahoo!**[35]

The Yahoo! search engine is used by the RNIB College during their Internet lessons. Subject One reported that size of the font was too small to comfortably access without moving close to the screen and physically moving his head to where the information could be read. The small icons at the top of the screen to locate 'What's New', 'Check Email', 'Personalize', and 'Help' were too small and the colours were not contrasting enough. The advertising banner placed above the search box was, unfortunately, the most clearly visible. The search box itself was not well highlighted. Yahoo! was not well received by the Subject.

Subject Two was frustrated with the lack of results. The font size was too small to access comfortably and text was cluttered too close together. There was insufficient contrast between hypertext link and background colour. The text was cluttered, but he found that the categories were listed alphabetically. This searcher could not identify the icons appearing at the top of the page without moving the magnifier to specifically view the images.

Subject Three received verbal output such as:

> 'links; links; comma link comma yahoo; comma link comma image comma; comma link comma options; comma link comma yahoo; mail dash free email account; comma link comma yellow pages dash; comma link comma people search; dash comma link comma maps; comma link comma classifieds; comma link comma personals; comma link comma chat dash; comma link comma emails; comma link comma calendar'.

He found the speech difficult to listen to, especially when the order of the speech appears to be random, e.g. yellow pages, then people search. He became visibly tired after listening to the speech output for thirty minutes. He became lost during the search and had no idea of where in the category search he had reached. The only way that was known to identify where the searcher was, was to backtrack using ALT + B until a page was found that could be identified.

Subject Four abandoned his search due to his frustration with the repetitive output. Continual linking through 'comma link comma Yellow Pages' etc. every time a new page was accessed was annoying, but could not be skipped over, as there was no indication where the actual results, i.e. Internet sites found, were located.

**Infoseek**[35,36]

Subject One's initial impression of the interface was of indifference. The font size was too small to comfortably access and it was suggested that the font colour could be brighter to aid easier readability. The advertising buttons were found to contain graphics that were too small to view comfortably. The overall impression was that it was the best of the three search engines that were accessed, but still had caused problems.

Subject Two said that the advertising icons were difficult to read. The hypertext links that were in bold font were far easier to read than the links on pages that were not emboldened. The Subject found this search engine to be the most difficult to access of the three that were selected.

Subject Three found that by the time a page was accessed, the speech relayed was the same. This information cannot be skipped over because there is no way of knowing where the new and relevant information begins. Some words were misheard and needed repeating, and also links such as: 'Click here to' followed by 'search Classical' did not make sense to the searcher. Infoseek also had 'without caption image' properties on their pages that were unhelpful and frustrated the searcher, as no information was gained, but time was wasted by having to access these links in order. The Subject did not feel comfortable with this search engine interface.

Subject Four reported that images without descriptions were unhelpful, and made him wonder if information was being missed out that other sighted searchers would find useful. The Subject did not understand the term 'Unsupported JavaScript' that was used. The searcher did not like the Infoseek interface.

### Other points raised by the Subjects

#### *Subject One*

The size of font used on Internet pages made the difference between an accessible site and a site that was difficult to access. Although having the most vision of the four Subjects overall, he preferred the font to be large enough to be viewed without having to put his nose within an inch of the screen. Grey background colour against a blue hypertext colour did not provide enough contrast. The use of continuous upper case text that was underlined proved difficult to read.

Many sites contain some form of advertising. This generally was found to be difficult to read due to the size of font used and the lack of definition on the graphics. Small animated graphics required close observation for several seconds before he could identify the image as advertising and not a valid part of the information search. As advertisements appeared within index listings as well as other parts of the page such as above the title of the page, there was no pattern that emerged to indicate which graphics could be discounted by their position. Subject One could not scan the whole page to discount advertisements, but had to access each one individually to assess their validity.

#### *Subject Two*

The main issues of accessibility were the size of font used and the colour contrast between the font and background. Grey background with blue hypertext was difficult to read. Where text was laid out in columnar form that did not make use of the screen width, the Subject was frequently required to relocate the mouse pointer with the scroll button. This was a tedious task due to him having to relocate the pointer icon to scroll down and then relocate his previous position in the text to continue reading. The Subject made good use of 'Related Stories' links once a relevant page was found. Unfortunately, the links were at the bottom of the page, so he was required to scroll down the entire page to be given this additional information.

The flashing of advertising text and graphics was uncomfortable for him as it distracted the eye from the text that was being read. Also when reading the advertising animations they continually changed and he could not read the text quickly enough. To determine whether pages were still loading, the Subject had to physically move to either the bottom left-hand corner or the top right-hand corner to gain feedback on the current status of the page.

#### *Subject Three*

The problems that were encountered by this Subject centred on the navigation from page to page and the interpretation of the speech synthesis to relate the page content. Pronunciation was often misunderstood and only by repeated access could the Subject begin to comprehend. The speech output disjointed sentences which broke the pattern of text. The repeated use of the words 'link' and 'comma' was tedious and required high levels of concentration. Image links without accompanying captions could not be described and he did not understand what 'image links' meant. The images and advertising graphics that had no available text could not be skipped over and were accessed sequentially to find any relevant text to the search. The Subject found this to be very time consuming.

On one particular site, the speech related all the indexed links, including advertisements without a caption, on the left-hand side of the page before the actual title and text content of the page. This proved to be tedious, as the Subject could not decide if the page was relevant or not until this information was given.

Although the same pages were accessed by Subjects Two and Three, when Subject Two retrieved relevant information by linking to 'Related Stories', Subject Three did not as he did not read down to the end of the page. Subject Three therefore missed out on possible relevant information.

### *Subject Four*

The use of Internet pages for commercial uses, i.e. selling goods, caused problems, for example when prices were given, there was no indication of whether the amount was in dollars or pounds. The synthesised speech was not always understood. The Subject reported that the use of 'gimmicky' words within a page did not translate well. For example the term 'Prev' meaning previous was not comprehended. Images and advertising without explanatory text were a waste of time. The Subject liked all aspects of a page to be explained as it all has to be accessed sequentially and parts cannot be skipped out. He stated that he preferred to have them described rather than an 'Image without caption' as he felt that there was a sense of 'missing out' on something that sighted users can see. Image links were described as 'Good' and 'Very Good' but this meant nothing to the Subject as they did not describe what the link actually did.

## Discussion and conclusions

The Web has the potential to open many avenues of resources and communication. There is a wealth of research underway regarding the evaluation of Internet search engines[37]. We hope this article will trigger a greater emphasis on evaluating such engines from the point of view of those with visual disabilities. With speech synthesis being used to access information, there is no longer a total reliance on Braille publications or pre-recorded tapes. This gives the user an independence from other people, which may be an important consideration to a blind or visually impaired individual.

Barriers that prevent access to the Internet include the interface design and the interpretation of speech synthesis to convey the content of the page. These obstacles can largely be overcome with a little forethought in the design process. Guidelines for accessible design and accessibility checkers are freely available on the Internet, but designers generally are more concerned with the 'look' of the page to sighted people than accessibility to a minority of users.

An accessible page does not have to be dull. It just requires more consideration in its design. Describing images in ALT text does not take long to do, but it enables a blind user to feel included in the content of the page. The term 'image' does not convey anything and may cause the user to feel excluded from some page content. Sites with a lot of graphics and sites with tables or frames could offer a text only version of their pages if they want to reach the widest audience possible. Advertisements within Internet sites are placed in prominent positions. Unfortunately, there is no pattern to this positioning from site to site, so visually impaired users need to take a close look to identify an advertisement and eliminate it from their information requirements.

At present, search engines are the primary way to search the Internet. Some search engines offer text only versions of their service. Unfortunately, the hypertext link function in AltaVista is sited at the bottom of the page in small font, which does not make it accessible. A better position could be at the top of the page in bold font, so the users who are most likely to require this facility can readily identify the link. The use of a bright colour around a search entry field in AltaVista was useful.

Subjects using speech synthesis could not identify the search entry field in Yahoo! or Infoseek. This may be due to the field being named a Data Entry field, which was not what they expected. Blind users have to listen to a lot of speech output to follow links. This requires concentration for long periods of time and is a tiring exercise. Speech synthesis did not tell the users which link had already been accessed. Users are therefore expected to remember where they have previously accessed. This is an unrealistic expectation.

The search criteria entered by visually impaired Subjects did retrieve some relevant information. The Subjects who were reliant on speech synthesis found retrieving relevant information much more difficult and time consuming. A simple explanation of how the search engine works and how results are displayed would greatly benefit all of the Subjects that took part, but especially the blind Subjects who became disoriented by the repetition of speech output on results pages. To enable visually impaired Subjects to scroll pages without having to relocate the scroll box each time, a mouse could be used with a scrolling wheel incorporated in the device so navigation would be at the user's fingertips.

While none of the sites visited were totally inaccessible to the Subjects, all had deficiencies. Layout is an important consideration; for example index links on the left-hand side of pages are relayed through speech synthesis before any content of the page. The title of a page appearing at the top is also helpful, as speech provides this information first so the user can determine whether it is worth continuing on that particular page.

An Internet access tool needs to be developed that would retrieve pages that have been designed with accessibility as a consideration. This could be easily implemented by embedding a <meta> tag with a keyword to indicate accessibility. As accessibility requirements may become part of the law and the designer's awareness is heightened, then more sites will be included and provide a useful service to visually impaired users. This facility could provide an indication of the links that have been accessed on the results page and therefore reduce the load on blind and visually impaired users to remember the sites that have already been visited.

Further research into speech synthesis software and the presentation of search engine interfaces, particularly the 'Results' page, for a visually impaired user is required. We only looked at three search engines. Further work on other search engines is clearly needed.

## References

1. Barrie, J.M. and Presti, D.E. 'The World Wide Web as an instruction tool. *Science* 274 (5286), 1996, pp.371–372.

2. *The making of a home page. http://www2.ite.net/edit.htm*

3. Pawan V. 'Human factors methodology for designing web sites.' In: Forsythe, C., Grose, E. and Ratner, J. (eds.) *Human factors and web development.* (Mahwah: Lawrence Erlbaum Associates Inc., 1998, p.153–172.)

4. *Frequently asked questions about blindness. http://www.rnib.org.uk/DOCS/wesupply/fctsheet/shortgu.htm#blindness*

5. *What can blind people see? http://www.rnib.org.uk/DOCS/wesupply/fctsheet/poster.htm*

6. *Population projections of blind and partially sighted people: 1996–2020. www.rnib.org.uk/wesupply/fctsheet/vipnos1.htm*

7. *Guidelines and advice on accessible IT systems and content. Resources for Visually Impaired Users of Electronic Library (REVIEL).* (Manchester: Manchester Metropolitan University: Centre for Research in Library and Information Management (CERLIM), 1998.)

8. *A brief introduction to disabilities. http://parallel.park.org/Guests/Trace/pavilion/populat.htm*

9. Laux, L. 'Designing web pages and applications for people with disabilities.' In: Forsythe, C., Grose, E. and Ratner, J. (eds.) *Human factors and web development.* (Mahwah: Lawrence Erlbaum Associates Inc., 1998, p.87–96.)

10. *Include Telematics Project. The World Wide Web accessibility. http://www.stakes.fi/include/accessib.html*

11. *The Internet and how to access it. http://www.rnib.org.uk/accessit.htm*

12. Harris, C. *Internet Magazine: level field. http://www.internet-magazine.com/features/disabled.htm*

13. Shneiderman, B. *Designing the user interface: strategies for effective human-computer interaction.* 3rd edition. (USA: Addison Wesley Longman Inc., 1998.)

14. Shneiderman, B. 'Designing information-abundant web sites: issues and recommendations.' *International Journal of Human-Computer Studies* 47(1), July 1997, pp.5–29.

15. *NY Times Cybertimes. http://www.rit.edu/~easi/law/weblaw1.htm* 12 January 1996.

16. Mynatt, E. 'Transforming graphical interfaces into auditory interfaces for blind users.' *Human-Computer Interaction*, 12(1&2), 1997, pp.7–45.

17. *Web Accessibility Initiative (WAI). http://www.w3.org/WAI/default.htm*

18. Thirty something (million): should they be exceptions? *http://parallel.park.org/Guests/Trace/pavilion/30_some.htm*

19. Waddell, C.D. *Applying the ADA to the Internet: a web accessibility standard. http://www.rit.edu/~easi/law/weblaw1.htm*

20. *Disability Discrimination Act 1995, Proposals for a Code of Practice, Rights of Access to Goods, Facilities, Services and Premises.* (National Disability Council, 1998.)

21. Oppenheim, C. *The legal and regulatory environment for electronic information.* 3rd edition. (Infonortics, 1998.)

22. Murnion, S. *A guide to Internet access for visually impaired people.* (Birmingham: Research Centre for the Education of the Visually Handicapped, 1996.)

23. Dixon, J.M. 'Levelling the road ahead: guidelines for the creation of WWW pages accessible to blind and visually handicapped users. *Library Hi Tech* 14(1), 1996, pp.65–68.

24. *Trace Center Overview. http://trace.wisc.edu/about/*

25. *RNIB. Hints for designing accessible web sites. http://www.rnib.org/wedo/research/hints.htm*

26. Fontaine, P. *Writing accessible HTML documents. http://www.yuri.org/webable/htmlcode.html*

27. *Microsoft accessibility and disabilities: for web page designers. http://www.microsoft.com/enable/dev/web.htm*

28. *CAST: Bobby. http://www.cast.org/bobby/*

29. *Welcome to Bobby. http://www.cast.org/bobby/*

30. *NCAM Web Access Symbol Now Available. http://www.wgbh.org/wgbh/pages/ncam/symbolwinner.html*

31. Graziani, P., Burzagli, L. and Tronconi, A. 'Internet for blind people: an information point.' *Colloque Inserm* 237, 1996. (New Technologies in the Education of Visually Handicapped), pp.291–295.

32. *Writing accessible HTML documents. http://www.itpolicy.gsa.gov/cita/wwwcode.htm*

33. Clarke, A. *Some UK websites for disabled people are totally inaccessible to them. http://www.abilitynet.co.uk/textonly/sponsors/features/feat3.htm*

34. *pwWebSpeak(tm) Chapters 1 to 5 Introduction. http://www.prodworks.com*

35. Maze, S., Moxley, D. and Smith, D.J. *Authoritative guide to web search engines.* (New York: Neal-Schuman Publishers Inc., 1997.)

36. *Search engine features chart. http://searchenginewatch.internet.com/webmasters/features.html*

37. Oppenheim, C., McKnight, C., Morris, A. and Lowley, S. 'The evaluation of WWW search engines: issues, problems and current developments.' *Journal of Documentation* 56(1), 2000. (accepted for publication)

## Acknowledgements

We would like to thank Dr. Ron Hinton, Department of Education, Loughborough University for his advice and support. We would also like to thank Kevin Connell, Principal of the RNIB Vocational College, Loughborough for his permission to access its facilities and to Denise Wain, lecturer at the RNIB Vocational College, for her valued input. Special thanks go to the four Subjects, without whom this project would not have been possible.

# Public libraries and EU information

*Mike Cooper*
*Reading Reference Library, Abbey Square, Reading RG1 3BQ*
*Email: Mike.Cooper@reading.gov.uk*

**Abstract** *Surveys the immediate background to, and establishment of, the European Commission supported European Public Information Centres (EPICs) operating in UK public libraries. Current structures for support and cooperation are outlined, and the author suggests that, although the network has been popular and successful, EPICs face challenges for the future, including external networking and internal cultural change.*

## Introduction

With the UK's Public Information Relay (PIR) network close to its fifth anniversary, and now renamed European Public Information Centres (EPIC), it is timely to take stock of the UK public libraries' involvement in European Union (EU) information.

The article will examine the background to public library involvement in EU information through the creation of the EPIC network. It will go on to look at the structures and functions of the PIR network, and end by looking briefly at some of the issues PIRs are facing as they are launched.

## PIR background – EU information and information policy

### *EU information networks in the UK*

The UK's Public Information Relay (PIRs) network is now the largest network of EU information providers in the UK, and conceivably in Europe. The network, which now includes the majority of the UK's public library authorities, was set up in 1994, to provide the general public with information on the EU and on their rights as EU citizens. At the time EU information provision in the UK was already relatively extensive, with networks serving small and rural business, academic research and, through the major deposit libraries, the general public.

To summarise, these networks were:

- *European Documentation Centres (EDCs)* – 40+ in number and based in major academic libraries. They received free either (a) all EU official publications or (b) selected EU titles, and were originally founded in the 1960s and early 1970s to support academic research into European integration. Support came from the Directorate General X (DGX) of the Commission. They developed strong links with one another and were the founders of the European Information Association. EDCs were obliged by contract to offer access to users from beyond their immediate academic community. In 1994 the EDC-DGX relationship was beginning to be re-worked, and these changes would lead to greater international networking and training.
- *Depository Libraries (DEPs)* – three major libraries received deposit collections of all EU publications. They were chiefly housed in major public libraries, serving diverse customer bases. DEPs generally function as, and alongside, EDCs.
- *Euro Info Centres (EICs)* – set up in the late 1980s by DGXXIII to offer information to Small and Medium-sized Enterprises (SMEs). Twenty plus units were housed in Chambers of Commerce or local development bodies. They offered commercial business information, and were supported by network conference and training. By 1994 some EICs were integrating into the newly established Business Links set up by the UK's Department of Trade and Industry, and were beginning to develop subject specialisms, e.g. environment.
- *Carrefours* – a small number of units set up to serve the needs of rural businesses. They functioned broadly as EICs.
- *Innovation Relay Centres* – providing information, advice and 'hand-holding' for businesses and academics seeking EU funding or trying to exploit the results of EU funded research and development.

- *Sectoral Relays* – then based chiefly on the Law Society and TUC serving their particular sectors.
- *Commission and European Parliament Information Offices* – principally those in London.
- *European Reference Centres (ERCs)* – small basic collections in some public libraries and FE/HE college libraries. Some of these were seen as the first stages towards the acquisition of larger collection status – paralleling EDCs, others were more or less dormant. Recently ERCs have either been closing or transmuting to become EDCs or EPICs.

In addition to these formally supported networks most trade associations and other professional bodies, together with many banks and other business advisors, provided EU information. The UK DTI was the main provider of information at government level, and most public libraries already offered some EU information as part of their overall provision[1].

### *PIR/EPIC rationale*

In principle, any of these information services were open to the general public, and the European Commission's own offices, and those of the European Parliament provided ready enquiry access. The arrival of EPICs has expanded the EU information network in the UK from around 70–80 units to potentially thousands, as it now embraces many, if not all, branch libraries in the country. As will be discussed later, this network represented a considerable commitment in terms of finance and effort by the European Commission in the UK and by the libraries concerned, and before considering the impact of the network, it is worth looking at the rationale behind its creation.

The 1999 European Parliament elections saw a low turnout. This happened against a background of public uncertainty about one of the most significant steps forward in European integration, the creation of a single currency, the Euro. EPICs began as a result of a not dissimilar climate resulting from another major step – the Maastricht Treaty that moved Europe from the European Community to the European Union.

Referenda in France and particularly Denmark, and background unease in other member states, notably the UK, combined with budgetary pressures to convince the Commission of a need to restructure their information providing activities. A survey by the EU's Study Group on Information and Communication Policy, headed by Commissioner Willi de Clerq, was that the message was fine, but it just was not getting through[2,3]. The 1993 De Clerq report[2,4] led to two major developments:

- a change in focus within the Commission's own offices to concentrate more on information for opinion formers;
- the decentralisation of information provision through information networks.

Public opinion polling[5] – and practical common sense – suggested that the most appropriate vector for information targeted at the general public was the public library. With a need to focus the work of the actual Representations on opinion formers, and with only one main office covering the UK, the London Representation could not both deal with general enquiries and carry out this role. Seen from the point of view of those looking at EU information provision public libraries offered a number of advantages:

- they were already part of the public's information seeking behaviour, and polls suggested that the public felt they should be offering more information on the EU;
- they already offered EU information;
- they offered multiple outlets at a local level;
- they would not require the costs in terms of establishment and start-up that a new network would;
- they offered existing information management and reference skills;
- by and large they had a tradition of networking together, and with, other providers;
- a series of seminars and conferences had shown that those running public libraries were keen to take on this role[3,6].

It should be added that the use of providers on an audience specific basis such as this fitted in well with the Commission's previous approaches to information provision. Rather than develop the roles of existing units – for example European Documentation Centres (EDCs) – the Commission had chosen to set up Euro Info Centres to provide information for small business and, whether by accident or design, had taken a 'horses for courses' approach. Although both the EDCs and EICs could be considered as possible contenders, neither network was wholly equipped, nor – from the point of view of a former member of both – wholly willing to take on a wider public information role. Not all EDCs, usually based in major academic libraries, could offer ready public access, and EICs had to balance the needs of commercial customers and financial survival. Both could – and still do – answer general enquiries when needed, but that was not and is not their designed role. The public library option was one particularly suited to the more developed and proactive library culture of the UK, and at the time the idea was unique to the Commission's UK Representation.

Although the context of the De Clerq report was perhaps almost overtly political (but not party political), and although it is possible to point to some underlying tensions over the exact role of the network, it is important to note that EPICs were set up to offer enhanced information provision, not to promote the idea of European Union. The founding belief was certainly that the EU was a fact, and that it should be supported, but in the UK, the working principle was that the public needed easier access to information on which to take its own decisions. It was up to each authority to decide how it interpreted its role.

Within a year the majority of public library authorities had joined the network, in England and Wales a national Expert – later Steering – Group had been set up and a Training Group was planning a programme of events. In Scotland the role was performed by a PIR User Group. The actual administration of the network – the organisation of conferences and training – is done by a private company, Relay Europe Limited.

By 1998 the picture of provision had changed slightly with the arrival of: European Resource Centres for Schools and Colleges, providing curriculum support material and advice; and Citizens Advice Bureaux, offering EU information in their particular context. The network as a result became much wider. Balancing this the Commission's London Representation had ceased to deal direct with the general public – signposting instead to PIRs – and ERCs had more or less passed from the scene.

What was, and is, particularly striking about the PIR initiative was that it was a genuine partnership of interests. EDCs when established had been a gift from the then EC to support academic research; EICs had been purely contractual. Now, although PIRs signed an agreement, the Commission was taking the role of one partner in a network, as facilitator rather than donor or client. The following section will look at the structures put in place, the bones of the PIR network.

## PIR structures[7]

### *Relationships with the Commission*

Each library joining the PIR network signs an agreement with the Commission's London Representation. The library agrees to house EU materials, provide staffing, promote the existence of the resource and to participate in network events and networking, reporting back to the Commission as needed.

In return PIRs are provided with:

- 50% discount on EU publications bought through The Stationery Office;
- start-up and promotional materials;
- guidance on possible stock;
- copies of free EU publications;
- a network newsletter;
- free places at network events;
- free places on national training events.

Of these, surveys suggest[8,9] that the discount and training are possibly the most popular aspects of membership.

Overall relations with the Commission's London Representation have proved excellent, with both parties recognising the constraints within which the other works. Importantly, the Commission has recognised that having decentralised it cannot dictate PIR operations. In fact, it is probably true to say that the UK Representation would not seek to do so. The PIR agreement is flexibly worded and its commitments far from onerous. Individual PIRs may interpret their role as their circumstances dictate. Most have appointed a given member of staff to have responsibility for the day to day operation of their service, usually as part of other reference and information based duties.

If there have been tensions, it has come chiefly as a result of PIRs' failure to understand the flexibility and hands-off nature of the Commission's role. It is arguable that librarians have not always recognised that, once the PIR agreement is signed, it is for them to determine what happens next and to match what they do to their local circumstances. Part of this may be due to the unfamiliarity of the subject area – a theme that will recur in this article – but part of it is also due to their unfamiliarity with dealing directly with a governmental body. It may also have something to do with the fact that it is not always easy for PIRs to see that they are only part of what a part of the European Commission does. As this article will go on to suggest, the nature of the subject being dealt with may also present some public libraries with a genuine cultural challenge.

### PIR resources

Relatively quickly it was realised that a broad brush approach to PIR provision, whilst generous might actually be overly generous, taxing the Commission resources and leading to inconsistency and a lack of co-ordination across the network. For both parties some consolidation and streamlining was necessary, if the network was to be effective. From 1997 on PIRs began to operate at three levels or 'Tiers'. These Tiers represent levels of provision and the basis of levels of service. Each level has a recommended level of stock, and a ceiling to its access to the Commission supported discount on publications.

- Tier 1 – six major libraries offering an in depth service. They will have a designated and dedicated EU information point, designated staff and usually a high level of public ICT access. These libraries offer a comprehensive and high level of EU information.
- Tier 2 – the bulk of town and county libraries. 'Tier 2s' or 'Level 2s' range in provision and resources from the very small up to those with virtually the resources of a Tier 1 – indeed there is open discussion of a 'Tier 2+' status at present. It is difficult to generalise, but on the whole they will offer a fairly comprehensive general EU service as part of their overall reference service, possibly referring out for EU law materials, although many are using the Internet or CD-ROMs to provide legal and legislative material.
- Tier 3 – smaller libraries and particularly branch libraries. Their role is primarily as a collecting point for enquiries, and to provide very basic information from a specially designed binder, provided to all network members by the Commission.

### PIR training

Although EICs were provided with supporting training, none of the library based EU information networks have received such comprehensive support. Given the number of libraries coming into contact with EU information, often for the first time, and given the proactive role implied, if not spelt out it in the EPIC agreement, staff expertise should prove to be a vital resource.

Most public librarians – indeed perhaps most librarians in the UK – have had little contact with EU information, as until quite recently this has featured little in library school curricula.

The EPIC training programme, devised by a voluntary training group, offered three free training 'Modules', joined later by a fourth. Each PIR could send two people on each course, and the courses were re-run as demand dictated. The Modules were:

Module 1 – a basic welcome to the network, and an introduction to the Commission's London Representation.

Module 2 – a fairly intensive session on the EU and handling EU publications.

Module 3 – rolling on from Module 2, dealing with sources for EU legislative material.

Module 4 – electronic sources.

Additional Modules 5 and 6 were piloted for all Relays.

In early 1998 it was decided that these Modules should be reorganised. Although they were almost universally popular, trainers had begun to find that the great blossoming of information on the Internet – in particular the arrival of new database products such as EUR-Lex, which offers basic free access to EU legal materials – meant that a hardcopy/electronic split was no longer viable or justified. Four Modules were arranged into two.

Module 1 – a basic introduction as before, but now including a little more 'meat' on the structure of the EU, and using 'homework' – basic level directed reading – to prepare trainees for the second Module.

Module 2 – a fairly intensive day looking at sources of EU information: legal, background and bibliographic. Practical sessions gave the chance to handle materials, and electronic sources were discussed alongside hardcopy alternatives.

These new Modules 'feel' more public library focussed that the old package and recognise the relatively low level of resources available to the great majority of EPICs. They are pitched at staff whose responsibilities include not only their EPIC, but also a range of other things, and who come from Tier 2 libraries, with moderate demand. As such much is condensed, and the focus is on effective, cheap solutions to actual demand, rather than on an overview of the possible world of EU information.

By 1998 the majority of EPIC members had had at least one, and usually two to four members of staff attend Modules 1–4, and at the time of writing the new-style Modules are underway. Both old and new formats have proved very popular, and it is arguable that the enhanced EU information handling skills that this has brought about has been one of the main achievements of the network. The reverse side of this is that it is forcing librarians from EDCs – who have provided the core of the pool of trainers – to re-evaluate existing tools, and learn about sources they would not have otherwise handled, hopefully spreading awareness both ways.

It is 'free' training that we are talking about, and a sceptic (or possibly a cynic) might argue that its popularity might owe something to this fact. The EPIC training means that it is possible for individuals to be given training at no cost at a time of shrinking budgets, and it may be worth considering the extent to which take up of this training reflects the nature of actual enquiries, which must be dealt with.

Discussion is underway about regionalising training, and work is going on on developing training packs and distance learning modules for Relay members. With regionalisation will come the need to pull non-librarians from other Relays into this training.

### *Network events*

From the outset it was clear that PIRs were conceived of as a network, and would be supported as such. The centrally planned training and information support offered has been discussed above, but generous support has also been given to networking events, both within EPICs and involving EPICs with other providers. These events have taken the following forms.

- *National conferences*
  At first cross-Relay planning events, but now true networking events involving EPICs with the members of other Relay networks. A national EPIC conference and Annual General Meeting is also organised, and fully supported by the Commission. This has attracted high level speakers, including Chris Smith, Heritage Secretary, and has proved an important component of the EPIC self-identity, as well as a valuable event in its own right.
- *Regional conferences*
  The importance of subsidiarity in the Commission's thinking about Relays, and practical common sense dictated that some of the most effective work done in 'networking' would be done at a regional basis. There have been problems in defining regions in a country that has only recently begun to think – in administrative terms at least – at a regional level. Again, regional conferences have happened at a cross-Relay level and within EPICs. Regions have now become more formalised, and in

general the lead for EPICs has tended to be taken by the Region's Level/Tier 1 EPIC, or by the most active Tier 2s. EPIC survey's show that a regional dimension is very popular, and the framework offers a great deal of potential for turning the less tangible benefits of networking into practical results in terms of referral networks, resource coordination and, by no means least, training.

- *Local events*
  Public libraries have been actively involved in local networking events for a long time. EPICs have provided another impetus for them to arrange formal or informal events to meet colleagues, discuss matters of common interest and compare experience. They have also offered a way to break unfamiliar regions into manageable units with a real identity 'on the ground'. In Reading we are part of a thriving local 'subregion' – the term is purely informal – originally set up under Berkshire County Council which includes the six unitary authorities of what was Berkshire, and our neighbours in Buckinghamshire, Milton Keynes and Oxfordshire. The focus has been on half-day basic training events for all staff to spread knowledge and break down barriers, and the events have attracted EPICs from Surrey and West London.

Network events have proved successful and popular, but it should not be forgotten that involvement does come with a price attached. If an EPIC is to be fully involved in its network, staff – and often one lead member of staff – can be sucked into a significant number of meetings and training events in a year. Practical issues aside, this does raise questions of cost benefit, and the extent to which personal professional development and the development of the service are one and the same. Many of the benefits of network participation are, on the surface, intangible, and it is easy for an EPIC to become a net donor to the network in terms of staff time and expertise. On the face of it this is not a problem, but in the real world of shrinking training budgets and pressure on headcounts it is clearly one which, from casual discussion with EPIC colleagues, many are having to face.

### *EPIC users*

To complete the EPIC picture it is necessary to describe their use and users. Unfortunately, this is quite difficult to do objectively, as full and fully compatible data are not yet available.

#### *Volume*

The PIR network was set up on the premise that, in the UK, at least, there was a latent, unfulfilled demand for EU information amongst the general – or to be more specific, voting – public, and that this demand would be best met by placing information where people said they would go to look for it: public libraries. It would be interesting to repeat the public opinion surveys on which some of the decision making behind EPICs was taken, as most EPICs feel that their demand is low and that, taking a recent EU survey as a guide[10], there is clearly unsatisfied demand. In this connection three questions need addressing.

- First and most obviously, is demand actually low? If so could it be higher?
- Second, if there is a gap between potential and realised demand, why is this?
- Third, how should EPICs react to this? Is it a problem? If so whose?

The third question will be dealt with later as a policy issue for the network as a whole.

Very obviously a great deal of research has to be undertaken before any firm conclusion can be drawn. Given the overall scope of this paper, all that we can do here is to broadly survey the issues and suggest areas for future research.

Direct evidence for the demand for EU information in public libraries comes from subjective impressions from colleagues, and from two self-assessment surveys conducted in 1997 and 1998 for the PIR network[7]. Table 1 compares these to the results of an internal survey from Berkshire and two external surveys[6,11,12]. As the base for these surveys was not always the same and their methods differed the data can only be regarded as being indicative. Work is underway to look at a more systematic and standardised form of data collection for PIRs, and this will include a definition of what is an 'EU enquiry' – something that has proved difficult to date.

*Table 1. Volume of demand for EU information.*

| Survey | Result |
|---|---|
| PIRs 1997 | Percentage of total enquiries reported varied from 0.1% up to 10%, averaging 3–4%. Numbers of enquiries stated were from 2–30 per week. |
| PIR 1998 | Average 45 enquiries per month. The most that was reported was that EU enquiries constituted 5% of total enquiries, for most the figure was less than 2%. |
| Berkshire 1996 | Covered 7 Reference Libraries was 0–2.5% of customers stating that they used EU information. Recent internal surveys at Reading tend to confirm c.2–3%. |
| Kelly and Nicholas 1996 | c.2% in one PIR. |
| Marcella *et al* 1996 | Unquantified – but felt to be rising with PIR membership. |

Overall then, it appears that EPICs find that not much more that 3% of all enquiries received are 'European' in content.

When, *Eurobarometer*, the public opinion survey run by the European Commission, asked if its respondents wanted more information on the EU they found that 23% of all respondents wanted more information on the EU[13]. In the UK, the figure was somewhat lower at 17%. If nothing else this suggests that a significant minority of a given public library's audience could be seen as the potential market for an EPIC. So why then are actual enquiries much lower than this? *Eurobarometer* suggested some possible explanations. Although the 1998 survey referred to here did not ask about use of libraries – just 'public information points' as a whole (it did show that most respondents found their information on the EU from television, and that most of those who would actually look 'proactively' for it would not necessarily think first of going in person to an information unit set up for the purpose[13].

### *Audience*

Public Reference libraries have a very wide customer base. The belief when EPICs were

*Table 2. EU information customers.*

| Survey | Result |
|---|---|
| PIRs 1997 | Demand pattern is College/School: Business: General Public in Ratio 6:2.5:1 |
| PIRs 1998 | Ranked as most important source of demand – Education: 36% General Public 34% Business 26% |
| Berkshire 1996 | Demand pattern is School/College: Business: General in Ratio 7:3:1 |
| Hertfordshire 1998 | Users say 47.4% Business/Work; 23.7% Personal Interest; 22% Educational Purposes[14]. |
| Kelly and Nicholas 1996 | Education: business: General, but not quantified. |
| Marcella *et al* 1996 | Ranked as 'significant' source – FE/HE: 66% School: 58% Business: 53% General 41% (69 if 'Jobseekers' included). |

*Table 3. Information demanded.*

| Survey | Result |
|---|---|
| PIRs 1998 | Type of material used ranked 1. Literature/Leaflets, 2. Statistics, 3. Business, 4. Law |
| Marcella *et al* 1996 | Topics becoming more popular with PIR membership were: 1. General EU activities, 2. Grants and loans, 3. Law |
| Hertfordshire 1996 | Users likely to use Geographical (55% responses), Legal (49%), Political (41%) |

set up was that this meant that they would be able to reach an unfulfilled demand from a general adult audience for general information about the EU and its policies. Tables 2 and 3 compare statements about the breakdown of the EU information audience in public libraries, a comparison that seems to suggest that the position is not this simple.

Realised demand seems to come from those who cannot ignore the EU: those in education who must study it and those in business who must work within its frameworks or who are seeking to benefit from it through business opportunities and funding. Certainly, this has been the experience of those providing EPIC training – those resources welcomed most by the audience are not those dealing primarily with 'citizenship' issues, but those that offer a mix of factual material for 'projects'. The great strength of the PIR network, of course, is that it can provide for all these needs. These interest groups are after all regular parts of any general public library customer base, although our own survey in Berkshire did suggest that EU users and general reference library users were a different set of people.

The 1996 Reference Library survey ran in the six main reference libraries of what was then Berkshire. Customers were asked to give their employment status, and this was then compared with data from previous overall library surveys for the County. The results showed a striking disparity between the evidence from reference libraries and that from the overall surveys. Table 4 shows the results from these, and Table 4a compares a ranking for the two communities[12].

To a certain extent this data confirms a basic picture of overall public library use and the background demographics of towns covered. However it is striking that those seeking work are twice as numerous in the reference library population, those in some form of education almost twice as well represented, and those actually working are similarly better represented than would have been predicted from the background library using population. Those 'looking after home' are markedly under-represented, as are retired people, although the latter remain an important group.

The inference here may be that reference library use is generated by an externally imposed need – the needs of work or study, or the needs of seeking work, whereas general library use is generated by a much more mixed range of needs. Clearly this is an area where more work is needed before firm plans for EPIC exploitation should be made, but it is interesting to note that the pattern here does resemble the 'work/study: everything else' model apparent in statements and findings about EU information use.

In this light, reference libraries were clearly the right place to put EU information, but perhaps not for the reasons planners thought.

The narrow view that EPICs serve the general public with general information on citizenship issues and that the specialist Relays, EICs and ERCs deal with business and education is clearly inaccurate. In all probability, it represents an overstatement of the simplicity of the initial concept of EPICs, but the view is one that must be addressed. One can argue that the breadth of PIR usership is a great strength for all Relays, and needs to be

*Table 4. Reference and General Library users from Berkshire surveys – by employment status.*

| Employment Status | Reference Libraries | All Libraries |
|---|---|---|
| In full/part time or self employment | 51.1% | 40.7% |
| Seeking work | 7.1% | 3.3% |
| Looking after home | 3.6% | 15.0% |
| School/college/university | 16.0% | 9.2% |
| Retired | 19.2% | 29.1% |

*Table 4a. Ranking by employment type.*

| Reference Libraries | All Libraries |
|---|---|
| 1. Full/part time/self employed | 1. Full/part time/self employed |
| 2. Retired | 2. Retired |
| 3. School/college/university | 3. Looking after home |
| 4. Seeking work | 4. School/college/university |
| 5. Looking after home | 5. Seeking work |

recognised – no one, neither the Commission nor the Relays themselves, will benefit from any form of compartmentalisation. Some services will be better delivered by certain providers, but unless a massive, (and, probably, foolhardy), campaign to re-educate the customer to suit the model is undertaken, then everyone must recognise the very great contribution public libraries have to make outside their intended field. For EPICs this may mean that they must in some cases adapt their resources and practices to suit the realised demand rather than match any preconceptions, and this must also be reflected in a more assertive stance adopted by network members. With Relays as a whole 'regionalising' it will be increasingly important that EPICs assert themselves as places that any customer can come to for EU information. They should also stress their skills in handling information in general terms, moving public libraries towards being an automatic first port of call for information.

## Policy issues

### *Demand – how 'low' is 'low'?*

Most PIRs would characterise their demand as 'low'. However it would be interesting to analyse PIR demand to find out the percentage represented by other subject areas. Put directly is 2% of all enquiries 'low' for a discreet subject area? A recent sample survey in Reading Library confirmed that about 2% of enquiries were specifically EU related[15]. However, other seemingly more mainstream topics showed relatively low demand – travel at about 3%, science at 1%, art at 4%, for example. UK law and legal or consumers' rights queries made up about 8% of demand on the day concerned, so it seems not unreasonable to wonder if 2% is that low after all?

With about 200–250 enquiries per day being received at Reading, it would appear about 20–30 people a week ask for EU information. Allowing for the fact that about two-thirds of customers never use the enquiry desk, but still use information, with around

700 visitors a day at Reading Reference Library (if customers statements are accurate) somewhere around 9–10 people a day are helping themselves to EU information. Given the unrepresentative nature of sampling, it is very difficult to be certain of these figures, but it could be that up to 40 people a week are using EU information at this one library. However, given a base population of about 120,000 people in Reading Borough, with possibly 17% (from the *Eurobarometer* survey) wanting more EU information we could have a potential market of up to 20,000 people.

It does seem that, even given an optimistic interpretation of library use figures, there is a clear difference between those saying that they want EU information and those actually using a public library for it. Balancing this Marcella *et al.*[6] found that most PIRs noted that demand for EU information had increased since they joined the network, and some PIRs are exceptionally busy with EU enquiries. It may be that simply by signposting an area as containing EU information, PIRs have encouraged users to make use of it. If this is so, then perhaps, regardless of the level of demand, they are succeeding in their aim. Certainly, given that a clear aim of the network was to relieve pressure on the Commission's London Representation any use of EPICs is a success. What with there being about 180 EPICs in the UK, this could mean that the Commission has been efficiently relieved of up to 1,800 calls per day. However, whilst this is quite possible, it does appear that either the potential demand is going somewhere else or that much of it is illusory – either in the minds of customers or in those of the providers.

Assuming that demand for EU information is low, and possibly that demand from the specific target audience is lower yet, what policy issues does this raise for PIRs? First, there must be the question of whether it is within the PIRs remit to act to increase demand. PIRs have an obligation to publicise their collections, but they are by no means agreed as to how far this should go in this respect. Some have campaigned actively in support of a range of EU related campaigns – the Euro, European Parliamentary elections. Others have felt that it is simply not their job to encourage their users to use one set of materials more, or have concluded that higher demand would not be sustainable. Given that 2% of enquiries is not disproportionately low, is this in fact really a problem. It would be interesting to examine why demand is considered to be low – or rather why librarians would, implicitly, like it to be higher. Clearly EU information is part of a rounded set of information on living and working in the UK, and should be offered as such, but given the likely take-up, how much should be offered?

Second, there is the question of resource allocation. Fortunately, with the 50% discount most of the materials necessary to run a basic EU service can be obtained fairly cheaply. However more and more is going on CD-ROMs, and more and more is going on the Internet. Is it legitimate to press for public Internet access for EU legal materials – from EUR-Lex on the Europa server, for instance – if only part of 2% of the customer base would use it or, more accurately, a part of it? Can expenditure on resources to cover for potential demand be justified, when possibly referral might be a better option for the few customers involved? If most customers could be dealt with by a Relay set up for that purpose, why shouldn't they be? More particularly, network participation whilst incurring no direct cost, perhaps, still incurs an opportunity cost. Training may be free, but will it be used? Certainly a training session can be a useful way of rewarding or motivating staff, and almost all training is worthwhile, but day-to-day library tasks must still be undertaken. From this point of view, a dedicated post – which would be one step to 'Tier 1 status' – would probably not be worth the effort for most Level 2 EPICs, unless demand increased dramatically.

### *Propaganda*[16]

If EPICs are to increase the demand for their EU collections they must promote them. Inevitably this will mean deciding what is being promoted – EU information or the EU. Contact with other EPICs suggest that some EPICs have found themselves the target of Euro-sceptical groups, and have had to deal with criticism, and requests to stock balancing material. This has made them wary of doing too much to 'push' their EPIC.

EPIC's relationship with EU information raises some interesting general questions about the authority and quality of information. Thus, for instance, the provision of UK Governmental information has not attracted the same degree of discussion or customer involvement, and yet UK governmental material is as much, or as little 'political' as EU material. There are two reasons why this is so. Firstly, librarians are not familiar with the EU, and so they feel that they are not happy in taking an assertive stance. Secondly, EU materials are relative newcomers, so they are less well understood, and have a less natural place in public reference libraries. Other than the actual issuing body, there is no real distinction between a Commission pamphlet on the Common Agricultural Policy and a UK Governmental booklet on Agriculture in the UK, and much less between the two sets of legislative material. What matters is that librarians are not always aware that there is little difference. To some extent the creation of dedicated 'European' areas may actually have contributed to this. UK official publications are not always kept separately, whereas EPICs are encouraged to create a separate area, and this may have resulted in compartmentalisation.

Eurosceptical opinion is only a real problem in that it reveals that, perhaps, many public libraries lack sufficiently well defined stock policies and information mission statements, so that they cannot readily justify a particular piece of information or readily refuse another. Alongside this, and perhaps more importantly, it reveals that public libraries have not always realised that their EU related stock can be thought of in exactly the same way as their other stock.

Clearly the more an EPIC provides information – as opposed to access to information – the more the content of the information it provides can be an issue. However with end-user searching of CD-ROMs and the Internet becoming common, EPICs need have no problems providing 'balance' if they believe that they do not already do so. The issues of access to information, and control of content that lie behind many Eurosceptical questions are important ones for librarians, they are ones which can be dealt with by clear overall information policies supported by the local authority.

### *Culture*

It is very easy to become bogged down in the detail of EU information. To generalise, as a library product, EU information (taken as a whole) is a very current commodity and changes frequently. It is also readily divisible by audience, level and format. All this tends to mean that strategies to delivery it effectively have to be based on information giving rather than collection building.

EU information collections have tended to be operated from reference libraries and often from the business section of a reference library. However, it is worth mentioning that, no matter how well developed the information giving culture is in a particular reference library, the outlook of public libraries tends, for historical, political and sound operational reasons, to be very much lending and book driven. EU information tends to be fact and answer driven. This is a sweeping generalisation, but it is one that highlights some aspects of why EU information has caused disquiet in some areas of the public library world. The nature of demand tends to be such that a 'stock' based response will not always work. Enquiries tend to be for specific facts or specific items of policy or legislative material – an EU collection is not enough, and resources must be devoted to retrieval tools and systems, rather than to reference books per se.

In other words, libraries providing EU information can be forced into giving answers when even their reference library culture may be based on tracing an item on the shelves, rather than seeking a particular fact for the customer. This type of work is highly staff and knowledge intensive, and presents an environment that perhaps resembles – albeit for the short time an enquiry happens – a law library or commercial information unit more that a traditional public reference library. This means for some it will be an inviting and exciting challenge, but it has to be recognised that not all librarians, and not all libraries, are prepared or able to embrace this type of work. It is easy then for tensions to arise between different aspects of the service, particularly when EU information is not seen as a traditional and well understood aspect of the service, and possibly where the information service provided had grown over the years rather than been planned.

If public libraries and librarians are uncomfortable with the information-giving role that comes with an EU information service then they may find that if they are not to be pushed into changing, then they must be careful to set out clear parameters for their service. Unfortunately, the involvement of external bodies, through the EPIC network, the Commission's own involvement, and through patterns of demand, may mean that they find an agenda being set for them. It has already been noted that potential users of EU information do not necessarily see public libraries as their first port of call. It may be equally true that local authorities and Government departments may need educating to see libraries as information providers; quoting one speaker at the 1999 EPIC annual conference – to see libraries as 'essential rather than just nice'[17]. Perhaps libraries need to consider what – on their own terms – being essential is all about.

This is not to paint a bleak picture, but public libraries may well have to think more rigorously than they have formerly done so. The outcome of this may be that they decide, or rather that some decide, that EU information is just an integral part of their overall service and cannot be allowed to set the pace. An alternative would be to use EU information to pioneer the way for other subject areas. In many ways it is well suited to this – it is important, can have a high impact and is in the news (for better or worse). Given an early commitment to electronic access it is also an area where information is available promptly and at many levels. In response to this some EPICs have promoted EU information extensively to local schools, forming strong links with European Resource Centres. Others have worked with their parent organisations, European officer or local Business Links and Euro Info Centres. In doing so they have helped to portray libraries as being something other than just a nice place to go to borrow books.

### *Technology*

Fairly clearly the EU sees electronic dissemination of its information as the way forward. In the long term, it may be that end user searching grows to such a point that there might be thought to be no reason to supply public libraries with hard copy materials. Equally clearly, this is a long way off.

Already key materials – such as the *Official Journal* – are becoming available only on CD-ROM or on the Internet. Whilst this is also true of UK Government information, the needs of distribution to 15 Member States have pushed the EU into offering more and more information electronically. Whilst public libraries are by no means cyber-phobic, there is a clear difference in resource allocation between buying a journal subscription and providing the hardware and support for a networked CD-ROM. It is important for all involved to recognise that at the present the apparently automatic substitutability of electronic tools for hard copy is more apparent than real for many libraries. The Commission cannot – and perhaps should not – afford to give every Tier 2 PIR at least one state-of-the-art PC with an Internet link. However whilst it cannot, someone must do so if the convenience and power of ICT to the producer is not to be wasted. Similarly, networking is greatly facilitated by email, and by the existence of common web sites such as Euroguide[18], but the benefits will not be enjoyed equally unless all can participate. As mentioned above, in times of tightening budgets, it is not always easy to argue that a given EPIC should have end user Internet access (probably free of charge) so that its customers may look at EU law online.

Staff awareness of and familiarity with ICT is also an important factor. Whilst training can be given, there must also be a willingness to tackle IT based tools and to deploy them seamlessly if their benefits are to be realised. Importantly too, if EPIC training is to be used the databases mentioned must be available and ready to use. At present contact with colleagues at training events has suggested that CD-ROMs continue to be popular tools, perhaps for just this reason – they are always there so long as the hardware works. However, as materials become available free of charge on the Internet, EPICs must look at making the Internet available with the same dependability as CD-ROMs, and with the same degree of staff familiarity and confidence in use. Obviously, EPICs are only a symptom of a changing Reference and Information culture here, but given their role and given the aspirations of the creators of their information product, they may find that they hit any technological threshold sooner than other subject areas.

ICT also has an important role to play in networking, in referral and in the administration of the network. Most public libraries now use email, and clearly this offers EPICs a quick and effective means to exchange information and to arrange events. Here, given basic access, issues of economies of scale – those discussed above as driving the producer side of EU information – can work in EPICs favour if resources permit. With low demand, and with scarce staff resources the prospect of someone else answering questions, albeit through online FAQs may be very attractive. However, at the risk of being flippant, for 'someone' to do this 'someone' has to be there. In this model, a lead library could find itself balancing out the economies of scale with extra staff time and involvement, apparently largely for the benefit of other people's customers.

Future development of an ICT driven EPIC network is attractive, but all partners must be prepared to confront the resource and conceptual issues this presents.

### *Networking – new partners/new paymasters?*

Traditionally the political structures within which public libraries have had to operate have consisted chiefly of their own local authorities, and whatever central government involvement, such as legal underpinning and statistical audits, were needed. For some libraries close contacts with TECs or local colleges and schools could pull other governmental bodies into the equation.

The creation of EPICs with direct European Commission support has added another player to this, and now with the establishment of the eight Regional Development Agencies (RDAs) the UK Government has added another 'team'[19,20]. The involvement of RDAs in issues of information infrastructure has yet to be finalised. Clearly, though, they must be looked to as a potential coordinator or at the very least a potential source of political support. If, as anticipated, European level cooperation and involvement with EU funding programmes is important to RDAs, then it is reasonable to expect that they will at least produce information on these activities, and that this information will need to be disseminated. It is, of course, unlikely that they will stop at that, and as they regionalise EPICs must be aware of these new, and potentially very powerful partners in their areas.

This is an issue faced by all public libraries, as they have not universally been part of previous regional economic development activities, such as local forums and the work of Business Links. Hopefully, RDAs will recognise that with appropriate input public libraries can serve any of the information audiences they may choose to reach. This is particularly true in a European context, but as has been mentioned EPICs and their parent libraries must seek to position themselves as active and professional information givers if this is to be so, and this must be matched by a recognition by the Commission and other network members of the skills they can bring and the diverse range of clients they serve. RDAs cannot be expected to share librarians' models of libraries.

As yet there are few indicators of the nature of the cultural base of RDAs in terms of information, but Government Offices, TECs and Business Links did not automatically look to public libraries as partners or even as vectors for their own information. Whatever their own policies it is clear that the arrival of RDAs, and a move towards the sort of regional structure more familiar to the EU from other member states will in all probability impact on EPICs. Certainly, with a well developed set of networks and an increasingly strong identity

EPICs have a legitimate claim to involvement, and it is important that this claim is supported both by their own authorities – and staff – and by the Relay network as a whole. You cannot sit back and expect RDAs to come to you.

## Conclusion

It is difficult for an insider with a commitment to Europe and European information to judge whether or not EPICs can be seen as a success. Disparities in demand, resources, institutional cultures and ambitions can make the network seem fragmented, and it is all too easy to dwell on too rigid a definition of what EPICs were created for. Looking objectively the EPIC network has:

- Established a new network of information providers with a strong logistical base and healthy infrastructure.
- Sustained that network and seen it grow to become the largest in the UK.
- Developed practical tools for use in even the smallest library.
- Brought a pragmatic understanding of EU information to more librarians – and other library staff then ever before.
- Kick-started successful regional and local groupings.
- Identified public libraries with active information provision for a powerful client.
- Maintained an open, frank and mutually beneficial dialogue with that client.
- Facilitated enhanced access to EU information for a very wide audience.
- Fostered related networks in other Member States.

Where issues remain to be resolved they are chiefly in the detail of the way EPICs work and in the future of the network. As such, the former are being dealt with, and are niggles – irritants – rather than threats. The latter represents a real challenge, but it is one where EPICs represent a facet of the picture of public libraries as a whole. To say that EPICs face challenges of changing demand, new technology, a need for internal cultural reassessment, and possibly redefinition and a need to define relationships with new political partners is only to state in particular the issues faced by public libraries as a whole.

## References

1. The overall provision of EU information in the UK is reviewed and contrasted with that in France and Germany in: Bender, P. 'Official information about EU affairs – co-operation networks of EU institutions and Member State Governments.' *European Access* 2, April 1999, pp.16–19.

2. Cooper, M. Editorial, *EIA Review* 2, July 1993 p.2–3.

3. Hopkins, M. 'National Consultative Conference, Stoke. Rochford. Report of the breakout session for information specialists.' *EIA Review* 2, July 1993, pp.4–13.

4. Group of experts on information and communication policy. *Reflection on information and communication policy of the European Community: Report by the group of experts.* Chaired by Mr W. De Clerq, MEP, March 1993. (Unpublished, Brussels, 1993)

5. European Commission. *Information deficit highlighted by EC survey.* Press Release, 10 November 1993. (London: European Commission, 1993.)

6. Marcella, R., Baxter G. and Parker, S. *The provision of information by public libraries in the UK.* (Aberdeen: Robert Gordon University, 1996, pp.1–3.)

7. Relay Europe Limited. *Results to the questionnaire 'The operation and future of the PIR'.* (Eastbourne: Relay Europe and PIR Steering Group, 1999.)

8. The overall structures are summarised in: European Commission. *FOLACL Expert Group on Public Information Relays Sub Group on Training: Public Information Relay Profile.* (London: European Commission, 1994.)

9. Public Information Relays: Steering Group. *Vision for the future: strategic action plan.* (Unpublished, 1996.)

10. European Commission. *Eurobarometer* 49, Spring 1998.

11. Kelly, A. and Nicholas, D. 'The public information relay: long term solution or half-way house.' *Aslib Proceedings*, 48(6), June 1996, pp.151–157.

12. Royal County of Berkshire Library and Information Service. *Reference Library Survey 1996* (Reading: Royal County of Berkshire Library and Information Service, 1996.)

13. European Commission. *Eurobarometer* 49, pp.3–7.

14. Hertfordshire County Council: Libraries Arts and Information. *European Information Needs.* Research report No.6 (Hatfield: Herfordshire County Council, 1996.)

15. Unpublished internal notes by the author based on a sample survey in September 1998.

16. Discussed more fully in: Cooper, M. 'Propaganda or fact? Dealing with Euro-sceptical material in public libraries.' *European Information* No.1, January 1999, pp.21–24.

17. Authors notes from 1999 National EPIC Conference, Birmingham.

18. Euroguide. *http://www.euroguide.org/*

19. Caborn, Rt. Hon. R. MP. 'RDAs and the new Europe: a government perspective.' *European Information Service* Issue 201, July 1999 p.4.

20. Powell, S. 'Working together: Government Offices and the regions.' *European Information Service* 201, July 1999, p.5–7.

## Note

Any views given here are strictly the author's own and do not represent the policies or views of his parent organisation – Reading Borough Libraries, the EPIC network or the European Information Association.

# Using cluster analysis to determine the media agenda

*Y.H. Kok and A. Goh*
*School of Applied Science, Nanyang Technological University, Singapore*
*Email: asegoh@ntu.edu.sg*

*D. A. Holaday*
*School of Communication Studies, Nanyang Technological University, Singapore*

**Abstract** *This paper describes a software tool that aids researchers in the study of agenda setting. Agenda setting theory claims that the mass media influences what the public thinks and talks about. The tool is used to cluster documents into topically coherent groupings that are to represent issues dominating press coverage. The documents are taken from the archives of online newspapers. In addition, the tool enables results to be visualised and displayed. Three methods were investigated for the purpose of clustering, of which the Group-Average-Linkage algorithm was chosen for the final testing. The choice of the clustering algorithm was predominantly made upon the quality of clusters produced. Comparisons between the computer-based results and a method involving human readers revealed comparable findings and potential usefulness of the software.*

## Introduction

The agenda setting function[1] purports that the mass media has an ability to influence public agenda through the events or issues communicated. Communication research into the correlation between media agenda and public agenda requires extensive use of manpower to read through daily news and to conduct interviews with the public[2]. Although there are statistical analysis tools[3] to assist in the research, the amount of pre-processing required before these tools can be used is still enormous. For this reason, a specially crafted software tool named Agenda was developed to aid in agenda setting research. The heart of the system consists of cluster-analysis algorithms that would enable agenda issues to be identified.

The clusters attempt to measure issues dominant in the media's agenda. This involves identifying issues emphasised in each article and then grouping these into clusters. Looked at over time, these clusters of documents represent one type of issue emphasis, measured in terms of the issue most emphasised in each article, which can be expected to peak and then be replaced as other issues come to dominate the media's agenda. Another type of issue emphasis has to do with how pervasively an issue is discussed throughout the newspaper, even in articles in which it is not the primary issue. Thus, another measure is the spread, which is the extent to which keywords identified by the algorithm during clustering turns up texts (newspaper articles) containing the keywords. This measure would be expected to yield a different pattern over time, lagging behind the aforementioned peaks but enduring longer.

Document clustering has an established history as a means for browsing large text collections. This development came about through the Cluster Hypothesis[4], which states that relevant documents tend to be more similar to each other than to non-relevant documents. Recent interest in document clustering has increased because of the information explosion from the Internet and the need to efficiently classify information[5].

Following this introduction, a brief description of the system architecture of Agenda is provided. Details of the clustering algorithms will then be discussed. The visualisation of the clusters is presented, followed by an evaluation of results before final conclusions are drawn.

## System architecture

Agenda aims to provide the user with an application for the analysis of the issues of the media agenda as characterised by news articles. It was designed with the following features:

- Ability to restrict relevant articles from a general archive using user-specified keyword(s).
- Provision of a summary of information of the articles, which includes headlines, dates and word counts.
- Grouping of the articles into topically coherent clusters, along with the generation of descriptive summaries formed by their weighted common words.
- Visual display of the clusters produced, together with their related information.
- Chart display showing the distribution of the clusters over time.

A system overview of Agenda is given in Figure 1. Agenda consists of four modules, namely, Article Search, Article Summary, Article Processing and Cluster Recognition & Representation. Article Search is used to restrict the articles through the use of search queries. The researcher may wish to perform specialised clustering on a selected subset of articles, for example those related to a particular topic such as foreign affairs or financial news instead of working on the entire news archive. The system is able to process documents in Hypertext Markup Language (HTML) format or text format. Next, the Article Summary pre-processes the articles to extract the headlines, dates, word counts and so on. Useful functions to view, save and print the articles are also provided. This is followed by Article Processing, which performs stopping and stemming and the generation of word frequency for each article. Stopping removes common function words from the articles, while stemming changes variations of words to their root forms. Up to this stage, the documents are only pre-processed for the main task of Cluster Recognition & Representation that performs clustering by grouping the articles into topically coherent clusters. The results of clustering and all the associated information are available for the user to view.

Agenda was developed using Microsoft Visual Basic 5.0 on a Pentium PC running Windows 95 Operating System. The compiled program is an executable application that can run on any 32-bit Windows platform.

## Document clustering algorithms

There are many clustering algorithms available[6,7]. Three techniques were investigated for performing clustering, namely, Similarity Score, Group-Average-Linkage (Cosine Measure) and Buckshot. The following sections explain the concepts involved, the methods employed and the associated advantages and disadvantages of each technique. A comparison of the three methods, as well as the justification for the eventual selection of Group-Average-Linkage for the final testing is discussed.

### *Similarity Score*

This method involves the computation of a Similarity Score between pairs of articles[8]. This score is taken to be an indication of the coherence of topic between the two documents, and acts as a basis for clustering them into a single entity using a single-link cluster structure. Every article is pre-processed to produce a list of word frequencies and weights. The weight of a word is dependent upon its frequency of occurrence and the mean word frequency in the article. It is an indication of the significance of the word in representing the article's content. The Similarity Score between two articles is taken to be the sum of the weights of all common words found between them. To qualify for grouping into a single cluster a minimum threshold is imposed on the score. Those articles with scores equal or above the threshold are deemed to be topically coherent and are grouped together. Documents are selected iteratively for comparison and fused to produce a list of clusters. The value of the threshold is set empirically, and can be changed by the user after observing the resultant clusters produced. A low threshold results in a small number of large document set clusters, while a high value produces numerous clusters with small document sets.

The main attraction of this method lies in its simplicity. Weights are summed up and compared with the threshold to determine clustering. Thus, the Similarity Score method

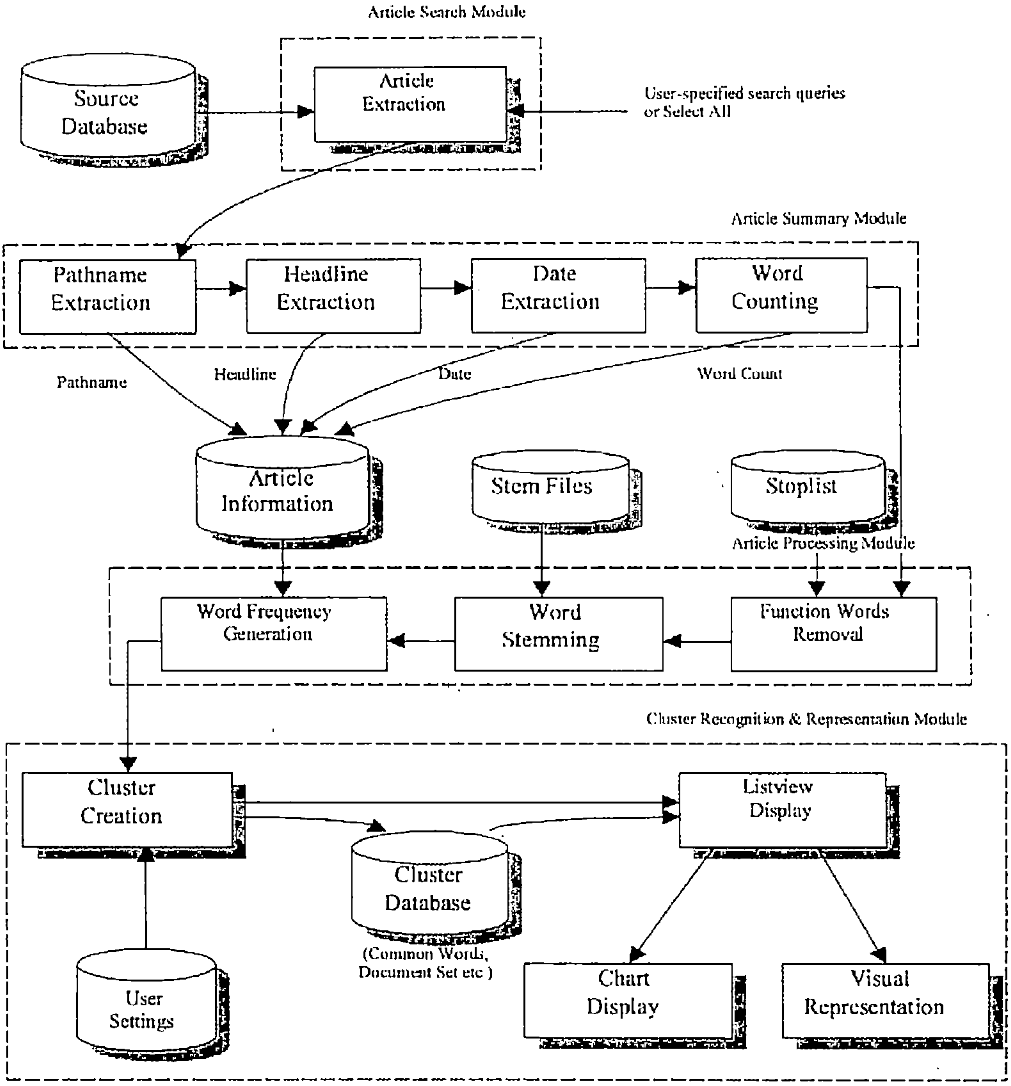


*Figure 1. An overview of Agenda.*

can be performed very quickly and enjoys a significant speed advantage over its counterparts. However, while the concept of Similarity Score is direct and logical, a closer examination reveals its potential shortcoming. The weight of a word, which is dependent upon the mean word frequency of the article it belongs to, suffers from a loss in information when it is summed with common words of the article being compared to. As the computation progresses, the significance of the word in the original document is further reduced. The absolute frequency of a word does not yield better results than the word frequency relative to the article's content. That was the reason why the weight was implemented in place of frequency. However, the subsequent summing up of weights, with no regard to the resultant word composition of the cluster, reduces the stated advantage.

Hence, it can be argued that while Similarity Score provides the benefits of simplicity and speed, the quality of clusters produced suffers with an increase in the number of articles.

***Group-Average-Linkage (Cosine Measure)***

In Cosine Measure, each document is treated as a 'vector' of length equal to the number of unique words. Each component of the vector has a value reflecting the occurrence and significance of the corresponding word in the document. The value assigned to the word components can be as simple as a binary scale, where the value of ones or zeroes represents the presence or absence of the word in the document. An absolute frequency of occurrence can also be employed. However, it would be desirable to use a value that incorporates the relationship to the document's word structure and length, and is a function of the word frequency as well. Hence, the weight is again utilised as the value assigned to the word component in the vectors for Cosine Measure. The information contained within these values, however, is better preserved with this method as compared to Similarity Score.

As with the Similarity Score method, a minimum threshold is set. This determines whether the two documents being compared are deemed to be topically similar. Extensive empirical testing has put this value to be approximately 0.25. The determination of the cutoff value is discussed at the end of this section. It should be noted that the threshold is dependent upon the application for which the measurement is used and the types of documents processed. For example, Cosine Measure employed in a search engine may have to use a different threshold from that employed by a content analysis application.

The algorithm produces vector norms results with a value between 0 and 1. The effect of this normalisation is to cluster documents with word frequencies having the same proximity of values. The degree of proximity depends on the minimum threshold set for clustering. To produce a set of words that will serve as the descriptors for the clusters, central words, namely those which have the highest weight in the group, are chosen.

The Cosine Measure enjoys significant advantage in terms of the quality of clusters produced as compared to the Similarity Score method. This benefit comes at the expense of a longer processing time, as computations are more complex. This is because each document is represented by a vector and the unique words it contains as components, resulting in large vector sizes.

***Buckshot***

The concept of the Buckshot algorithm employs the use of Cosine Measure for similarity computations between documents, but differs from Group-Average-Linkage in the way documents are iteratively compared. The algorithm can be stated simplistically as follows:

- Find *k* centres, where *k* depends on the number of documents to be processed.
- Assign each document in the collection to a centre based on its similarity.
- Refine the partition so constructed.

*k* documents are randomly chosen as the centres for comparison. Once *k* centres have been found, each document in the collection must be assigned to one of those centres based on the Cosine Measure criterion. This is done by comparing each document with all the centres, and assigned to the one with which it has the highest Cosine Measure. After an initial clustering, refinement is carried out. Two additional algorithms, Split and Join, are used for this purpose. The Split algorithm separates poorly defined clusters into well-separated parts and Join merges clusters that are too similar. Both algorithms use Cosine Measure as the determining criterion as well.

***Comparison***

After extensive testing with collections of news articles from the various sources, the comparison between the three clustering techniques are summarised in Table 1. From the testing results, the following observations were made.

- *Cluster quality produced*
  The clusters produced by Group-Average-Linkage and Buckshot were significantly superior to those obtained using Similarity Score. The disparity in cluster quality became increasingly evident as the number

*Table 1. Comparison between various clustering techniques.*

| | **Similarity Score** | **Group-Average-Linkage (Cosine Measure)** | **Buckshot** |
|---|---|---|---|
| Advantages | • Simple computation<br>• Fast clustering | • Good quality clusters produced | • Does not increase at the same rate as increases in database size |
| Disadvantages | • Does not take into account frequency and proportion information<br>• Quality of clusters deteriorates as size of database increases | • More complex computations<br>• Relatively longer clustering time | • Initial choice of *k* centres chosen randomly<br>• Non-deterministic |

of test articles increased. This is expected as the detrimental effects of non-proportional scoring and assignment becomes more conspicuous with an increase in cumulative processing.

- *Computational time*
  Similarity Score had the best performance in terms of processing time. The interesting point observed was that, while Buckshot promises a faster processing time over Group-Average-Linkage the actual measured processing times for both algorithms were surprisingly similar. One possible explanation could be that Buckshot was first proposed and tested on a collection size of documents in the range of thousands, whereas a relatively small set of over two hundred test articles used in our test may not be significant to reveal its processing time advantage over Group-Average-Linkage.
- *Deterministic nature*
  Since Buckshot and Group-Average-Linkage both produce clusters of similar quality and have comparable processing times, the deterministic nature of the latter becomes an important factor to take note of. Buckshot requires the random selection of *k* centres prior to grouping the articles. This means that with every new operation on the same collection one may obtain slightly different clusters. In other words, the results are non-deterministic and, very often, non-repeatable. In view of the user-requirements for Agenda this is an undesirable effect.

### *Determining control parameters of Group-Average-Linkage*

The cutoff value used for the Cosine Measure comparison was set at 0.25. This empirical setting was based on a series of experiments with different values to observe the effect of changes in the threshold. As shown in Figure 2, the number of multi-document clusters increased as the cutoff value was lowered. This means that with less stringent criteria for clustering, more documents were grouped as topically coherent. A lower cutoff value of 0.1 results in 99 per cent of the daily news articles being clustered. This, however, does not mean that all clusters enjoyed a high degree of relevancy.

The quality of the clusters produced with different cutoff values was also examined in order to yield an appropriate default setting for Agenda. As illustrated in Figure 3, a threshold of 0.4 to 0.5 produced clusters with a high level of content relevancy. Values between 0.2 and 0.3 resulted in a small, though negligible number of inappropriately grouped articles. This number, however, increased radically as the cutoff value was brought down to 0.1. In other words, a cutoff of 0.1 produced many incoherent clusters. From the findings, a default value of 0.25 was deemed appropriate to produce a substantial number of cluster sets with reasonably good quality.

*Figure 2. Effect of cutoff values on documents clustered and cluster sets.*

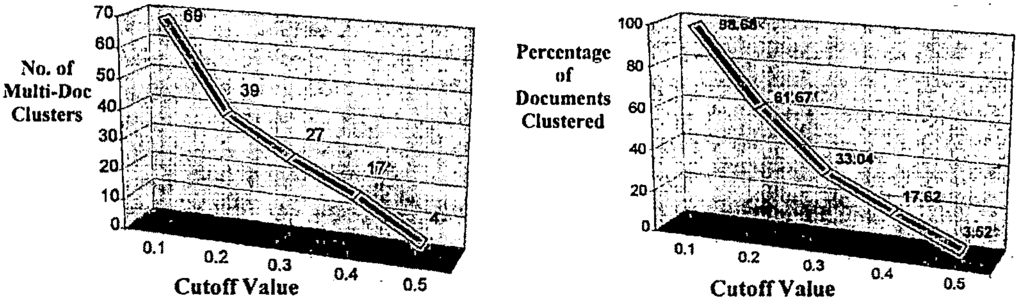
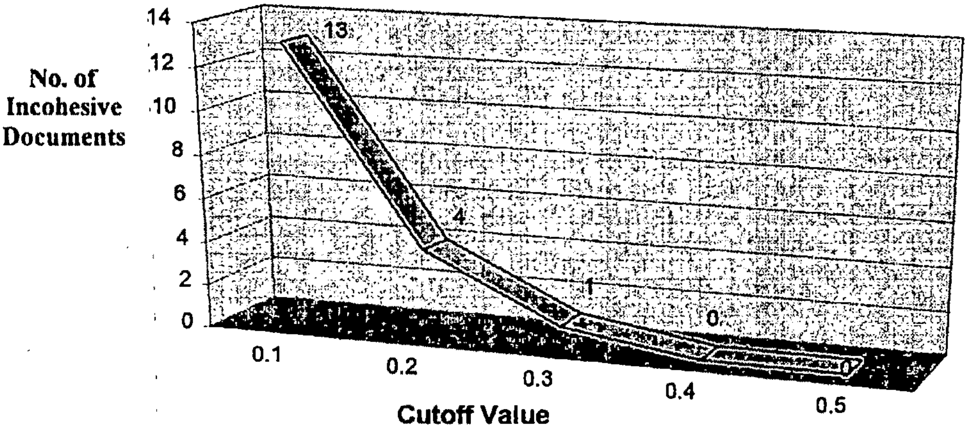


*Figure 3. Effect of cutoff value on cluster cohesiveness.*



**Visualisation**

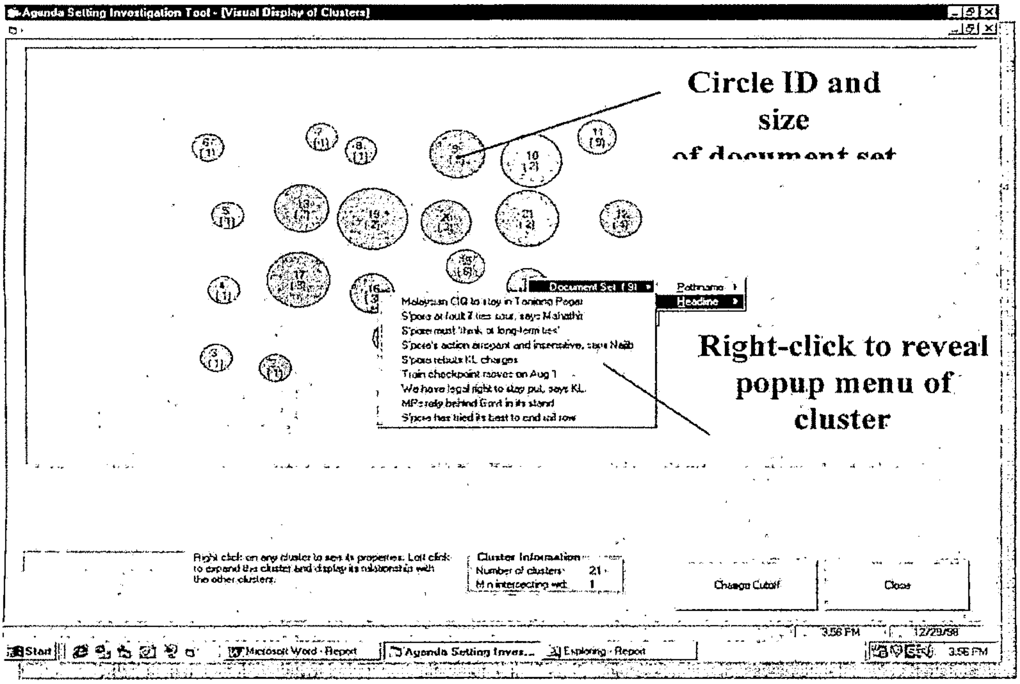
The visual representation of the clusters offers a more user-friendly graphical interface for users to investigate the results. Clusters are represented as circles, the size being a function of the number of common words contained within the cluster. Each circle is identified by a unique number and the size of its document set is shown in brackets below the number. Pop-up menus are used to display a cluster's properties. Information available includes location and headlines of the document set, the descriptors for the cluster, and its intersecting relationship with the other clusters. Figure 4 shows a pop-up menu in the visual display.

Information common to the clusters can be further investigated through another detailed display as shown in Figure 5. This display is presented when any of the circles in Figure 4 are selected. Seemingly different clusters may be related to one another through similar issues reported in both of them. The provision of this display helps to identify these links. The user can choose any of the other circles to compare with the selected cluster. Results of the comparison, namely the common words between these two circles, are displayed at the bottom right panel of the form.

***Timeline display***

The purpose of this display is to plot the distribution of the clusters over time. This is accomplished by plotting an XY scatter graph[9] of the word count of the articles against the dates when the articles appeared. This timeline display reveals temporal information while word count acts as an indication of the prominence of the article.

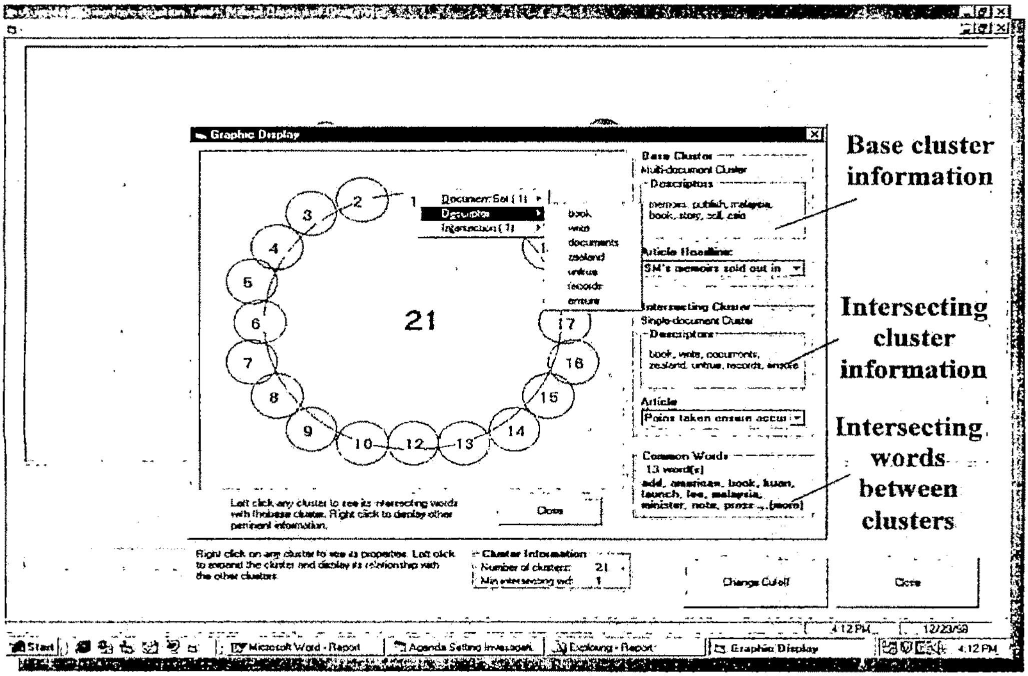
*Figure 4. Cluster visualisation with pop-up details.*



Malaysian CIQ to stay in Tanjong Pagar
S'pore at fault 7 ties sour, says Mahathir
S'pore must 'think of long-term ties'
S'pore's action arrogant and insensitive, says Najib
S'pore rebuts KL charges
Train checkpoint moves on Aug 1
We have legal right to stay put, says KL
MPs rally behind Govt in its stand
S'pore has tried its best to end rail row

Right click on any cluster to see its properties. Left click to expand the cluster and display its relationship with the other clusters.

*Figure 5. Details of specific clusters.*



memoirs, publish, malaysia, book, story, sell, asia

SM's memoirs sold out in

book, write, documents, zealand, untrue, records, ensure

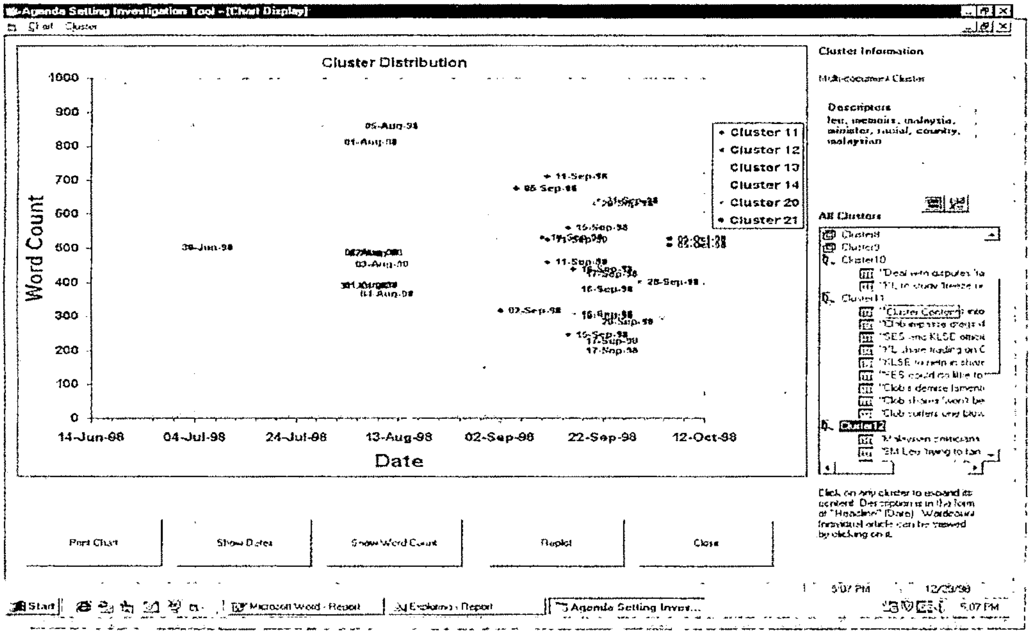
Points taken ensure accu

13 word(s)
add, american, book, kuan, launch, lee, malaysia, minister, note, prize ... [more]

Left click any cluster to see its intersecting words with the base cluster. Right click to display other pertinent information.

Right click on any cluster to see its properties. Left click to expand the cluster and display its relationship with the other clusters.

*Figure 6. Timeline chart.*



This facility is a particularly useful feature for investigating the spread or concentration of issues over time. For example, researchers may discover certain topics surfaced or were featured prominently during a particular period. The issues do not have to be concentrated in any particular 'peaks' in time to yield useful analysis. Some issues may be discussed pervasively during a certain period but still continue to be mentioned, albeit with a lower frequency, in the ensuing time periods after the 'peaks'. A maximum number of six clusters can be plotted at any one time on the chart. This user-definable limit is imposed to ensure the chart is not overloaded with clusters, resulting in an unclear presentation. Upon selection of the clusters to view, the chart is plotted and displayed. Figure 6 shows a sample of a timeline chart for selected clusters.

Each cluster is assigned a unique symbol and is colour-coded to facilitate clearer visualisation. A Replot function is provided to redisplay a different set of clusters. For easy reference to the clusters' content and information a treeview is provided in the chart. Each entry in the hierarchical listing is represented by a short description, which can be expanded to reveal more information about the document.

## System evaluation

This section focuses on the comparison between the results produced by Agenda and an Intelligent Reader Method. In this Intelligent Reader Method readers do content analysis of newspaper articles over a specified period of time. A coding sheet containing categories identified by the researchers is given to readers. Examples of these categories include Cost of Living, Election, World Affairs etc. Upon reading the daily newspaper, the reader is expected to provide a value for each of the categories, reflecting the level of attention given to each category that day. The values for this 'level of attention' vary from 0 (no attention at all) to 5, the latter showing that the category had been of outstanding concern that day. The overall emphasis was averaged for each week using double coding for reliability. The news articles used for testing and analysis were taken from a Singapore newspaper, the *Straits Times*, for the period between 8th and 14th February 1996. These

*Table 2. Summary of test results produced by Agenda.*

| **Summary of Results** | |
|---|---|
| No. of Multi-Document Clusters | 31 |
| No. of Single-Document Clusters | 128 |
| Average Cluster Size | 3.194 |
| Total No. of Documents Processed | 227 |
| Total No. of Documents Clustered | 99 |
| Percentage of Multi-Document Clusters | 43.6% |
| Percentage of Single Document Clusters | 56.4% |
| Cosine Measure Cutoff Value Used | 0.25 |

articles were obtained from the archives of Newslink[10] and Reuters[11]. A total of 227 articles were culled from daily issues during the test period. Instead of analysing the entire newspaper, articles were taken from pages one to three, the Home section for local news and letters to the editor. No further search restriction was imposed on these articles. The length of these articles range from 57 to 1,710 words with the average being 407 words.

In order to enable comparisons to be made with the results produced by the Intelligent Reader, user-defined categories identical to the coding sheets were set up. Each category is accompanied by user-defined keywords and cluster conditions, enabling the tool to group documents to any of the categories defined. These categories are saved as templates, thus facilitating re-use in future studies. Furthermore, fine-tuning of categories can be done by the researchers depending on the scope of the study.

A summary of results produced by Agenda for the test period mentioned above is given in Table 2. Details of a typical cluster are shown in Table 3.

The coding sheets produced by the 'Intelligent Readers' are collated to form a list of issues deemed as prominent ones raised by the press during the test period. On the other hand, the number of documents in each cluster is seen as an indication of issue emphasis. Thus, two lists of categories produced by Agenda and the Intelligent Reader respectively were formed. The lists are in descending order of issue importance as stated by the human readers.

The top and bottom five categories determined by Agenda and the Intelligent Reader are given in Table 4. There is a high level of similarity between the two methods in terms of measuring issue emphasis. The issues with high emphasis include Government Policies, World Affairs and the Election and were detected by both approaches. At the other end of the spectrum, the lack of emphasis given to issues such as Poverty, the General Economy, Social Class, Media and Foreign Interference were equally apparent. A conclusion was drawn that Agenda has potential in its ability to detect media emphasis.

## Conclusion

A tool to assist in identifying media agenda has been developed. The goal is to alleviate the tedium of employing many professional readers, thereby reducing the time and cost of identifying media agenda. The results show that such a tool holds promise though further enhancements are desired. Furthermore, the tool has the flexibility to allow researchers to do further investigation into document content.

*Table 3. A summary of a typical cluster.*

| | |
|---|---|
| Cluster No. | 137 |
| Size of Document Set | 11 |
| Total No. of Common Words | 5 |
| Similarity Score | 0.345 |
| Descriptors | China, chinese, president, report, taiwan |
| Documents' Headlines | ● Taiwan's Lee Reiterates Stand on Achieving Democracy<br>● China to Taiwan Panel – Don't Split Motherland<br>● US Warned Against Intervention<br>● China Will Act if Taiwan Leaders Invited to Games<br>● French Fears on China-Taiwan Tensions Eased<br>● Taiwan Leader 'Poured Heart Out' to Japanese Novelist<br>● Olympic Panel Says It Will Not Invite Taiwan Leaders to Games<br>● Taiwan Confirms China's Ground Forces Mobilised<br>● Taiwan, US 'Won't Allow Conflict With China'<br>● Taiwan's Military Goes on the Alert<br>● Taiwan Sets Up Panel To Ease Tension With China |

*Table 4. Comparison of categories ranking between 'Intelligent Reader' and Agenda.*

| | **Intelligent Reader** | **Agenda** |
|---|---|---|
| Top five categories | ● Government Policies<br>● World Affairs<br>● Election, voting and politics<br>● Housing & Upgrading<br>● Health | ● World Affairs<br>● Personal Wealth and Materialistic concerns<br>● Election, voting and politics<br>● Government Policies<br>● Cost of living; People (tied) |
| Bottom five categories | ● Poverty and subsistence of living;<br>● General economy/productivity;<br>● Social class, inequality;<br>● Media issues;<br>● Foreign interference in domestic issues | ● Ethnicity, race and chauvinism;<br>● Social class, inequality;<br>● Workers' rights/unions;<br>● Media issues;<br>● Foreign interference in domestic issues |

Even though the Group-Average-Linkage (Cosine Measure) algorithm has produced satisfactory results, it suffers from two major drawbacks. The first is the huge vector space involved. This space requirement imposes a severe computational constraint on the document processing speed. As the document collection increases, so does the time to compute similarities. In fact, many text-processing systems have already tried to move away from the vector space representation. Some of the common methods include recasting the vector space into one of much lower dimensionality and greater orthogonality through a kind of factor analysis[12]. Secondly, there is an absence of semantic reality. That is to say, there is insufficient emphasis placed on the context of a word. The presence or absence of a word may contribute to a 'more accurate' classification. One possible method to overcome this drawback is the inclusion of a thesaurus. The strong assumptions of orthonormality by Cosine Measure can be alleviated through a comparison based on multiple meanings of words rather than simple textual comparison[13].

## References

1. McCombs, M. E., and Shaw, D.L. 'The evolution of agenda-setting research: twenty-five years in the marketplace of ideas.' *Journal of Communication* 43, 1993, pp.58–67.

2. Holaday, D.A. and Kuo, E.C.Y. 'Upsetting the agenda: media and the 1991 Singapore election.' *Gazette* 51, 1993, pp.197–218.

3. SPSS Advanced Models (formerly known as Advanced Statistics) *http://www.spss.com/software/spss/*

4. Rijsbergen, C.J. Van, *Information Retrieval.* 2nd edition (London: Butterworths, 1979).

5. Zamir, O. and Etzioni, O. 'Web document clustering: a feasibility demonstration.' In: *21st Annual International ACM SIGIR Conference on R&D in IR*, 1998, pp.46–54.

6. Fraley, C. and Raftery, A.E. 'How many clusters? Which clustering methods? Answers via model-based cluster analysis.' *The Computer Journal* 41(8), 1998, pp.578–588.

7. Willett, P. 'Recent trends in hierarchic document clustering: a critical review.' *Information Processing & Management* 24(5), 1988, pp.577–597.

8. Harman, D. 'An experimental study of factors important in document ranking.' In: *9th Annual International ACM SIGIR Conference on R&D in IR*, 1986, pp.186–193.

9. Cutting, D.R., Karger, D.R., Pedersen, J.O. and Tukey, J.W. 'Scatter/gather: a cluster based approach to browsing large document collections.' In: *15th Annual International ACM SIGIR Conference on R&D in IR*, 1992, pp.318–329.

10. AJR NEWSLINK: *http://ajr.newslink.org/nonusasin.html.*

11. *Reuters Asian News: http://straitstimes.asia1.com.sg.*

12. Schutze, H. and Silverstein, C. 'Projections for efficient document clustering.' *20th Annual International ACM SIGIR Conference on R&D in IR*, 1997, pp.74–81.

13. Cohen, W. and Singer, Y. 'Context-sensitive learning methods for text categorization.' *19th Annual International ACM SIGIR Conference on R&D in IR*, 1996.

**Subscriptions**

*Aslib Proceedings* is published ten times per annum and supplied to corporate members of Aslib. For 1999 the annual subscription rates are £170/$280 (£136/$224 to members of Aslib). For subscriptions and back issues please contact: Portland Press Ltd, Commerce Way, Whitehall Industrial Estate, Colchester CO2 8HP, United Kingdom. Tel: +44 (0) 1206 796 351. Fax: +44 (0) 1206 799 331.

Aslib, The Association for Information Management, is a world class corporate membership organisation with over 2000 members in some 70 countries. Aslib actively promotes best practice in the management of information resources. It lobbies on all aspects of the management of, and legislation concerning, information at local, national and international levels. Aslib provides consultancy and information services, professional development training, conferences, specialist recruitment, Internet products, and publishes primary and secondary journals, conference proceedings, directories and monographs.

Further information can be obtained from: Aslib, The Association for Information Management, Staple Hall, Stone House Court, London EC3A 7PB, UK. Tel: +44 (0) 20 7903 0000. Fax: +44 (0) 20 7903 0011. Email: aslib@aslib.co.uk *http://www.aslib.co.uk/*

Periodicals Postage Paid at Rahway NJ and at additional mailing offices.
**POSTMASTER:** Send address changes to PROCEEDINGS, c/o Mercury International, 365 Blair Road, Avenel NJ 07001

© 1999 Aslib and contributors ISSN 0001 253X